

OM of: US-09-303-518D-131 to: PIR_71:* out_format : pfs

Date: Jun 30, 2002 7:42 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-Q/Cgnt2_1/USPTO.spool/US09303518/runat_28062002_142713_4317/app_query.fasta_1.23501
-DB=PIR_71 -OPMT=fastan -SUFFIX=trp -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPEXT=0.000 -GAPEXT=0.000 -GAPOP=4.500
-GAPEXT=0.050 -GAPOP=10.000 -GAPEXT=0.500 -GAPOP=6.000
-DELEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-FLEX=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=100 -MODEALIGN=200 -THR_SCORE=oct -THR_MAX=100 -THR_MIN=0
-ALIGN=45 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZ=500
-MINLEN=0 -MAXLEN=200000000 -USER=US09303518 -CGNL_1_1092
-NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
-NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-303-518D-131

Query length: 1344

Database: PIR_71.*

Database sequences: 283138

Database length: 96089334

Search time (sec): 504.980000

score_list:

Sequence	Strd Orig	Zscore	EScore	Len	Documentation
pir2:A81919	+ 2222.00	3060.25	5.1e-163	447	i probable sodium-translocating N
pir2:D81185	+ 2221.00	3058.87	6.1e-163	447	i sodium-translocating NADH dehyd
pir2:I64002	+ 1646.00	2264.12	1.1e-118	447	i sodium-translocating NADH dehyd
pir2:AG0393	+ 1600.00	2260.54	3.9e-115	447	i NADH dehydrogenase (ubiquinone)
pir2:S51015	+ 1488.50	2240.45	1.5e-106	465	i sodium-translocating NADH dehyd
pir2:G82034	+ 1476.50	2029.52	1.3e-105	462	i sodium-translocating NADH dehyd
pir2:G83272	+ 1358.00	1866.10	1.7e-96	445	i sodium-translocating NADH dehyd
pir2:C81751	+ 453.00	614.80	8.0e-27	465	i probable sodium-translocating N
pir2:D86583	+ 450.50	611.31	1.3e-26	467	i ubiquinone oxidoreductase, alph
pir2:F72040	+ 450.50	611.31	1.3e-26	467	i probable sodium-translocating N
pir2:F71489	+ 432.00	585.78	3.3e-25	465	i sodium-translocating NADH dehyd
pir2:B82252	+ 172.00	221.11	4.0e-05	801	i RnfC-related protein VC1015 [im
pir2:R80692	+ 151.50	194.48	0.0014	673	i probable NADH reducing dehydrog
pir2:S59310	+ 143.00	177.90	0.0073	1104	i probable membrane protein YMR3
pir2:B64136	+ 135.50	170.42	0.0118	1791	i hypothetical protein KIAA0324
pir2:D84308	+ 132.00	166.16	0.0242	451	i rnfC protein homolog - Haemophi
pir2:A80670	+ 131.00	169.58	0.0498	473	i hypothetical protein TM0244 -
pir2:D83903	+ 129.50	158.44	0.0820	1199	i probable ferredoxin PA3491 [imp
pir2:S39893	+ 128.00	152.40	0.1185	1802	i membrane protein [imported] -
pir2:T43481	+ 126.50	162.45	0.1132	519	i HKR1 protein precursor - yeast
pir2:T39903	+ 121.50	154.46	0.2824	580	i rnfC protein - Rhodospirillum rub
pir2:F85769	+ 121.00	154.57	0.3023	534	i probable mucin DKFZp434C196.1
pir2:S90921	+ 121.00	151.39	0.3280	740	i serine-rich protein - fission y
pir2:T52257	+ 121.00	150.98	0.3315	772	i probable membrane protein Z2636
pir2:T05352	+ 119.00	150.18	0.4493	631	i probable membrane protein [impo
pir2:T64919	+ 118.00	145.82	0.5791	857	i episialin - mouse
pir2:T47141	+ 117.00	145.87	0.6665	740	i hypothetical protein F8B4.120 -
pir2:T34513	+ 116.00	147.77	0.7314	528	i probable iron-sulfur protein bl
pir2:A45294	+ 115.00	127.94	1.40	3507	i gastric mucin (clone PGM-2A) -
pir2:T43311	+ 114.50	142.29	1.04	749	i hypothetical protein ZK783.1 -
pir2:T39207	+ 114.50	132.37	1.34	2073	i balbiani ring 2.1 - midge (Chir
pir2:F75196	+ 114.50	132.37	1.34	2073	i fatty acyl-CoA synthase (EC 2.
pir2:E86185	+ 113.00	139.35	1.39	819	i pyruvate,water dikinase, subunit
pir2:T86195	+ 112.50	145.59	1.27	402	i pyruvate,water dikinase (EC 2.7
pir2:A36904	+ 112.50	144.15	1.32	466	i hypothetical protein [imported]
pir2:T43023	+ 112.00	144.83	1.39	466	i protochlorophyllide reductase
pir2:T30635	+ 112.00	141.35	1.52	579	i ceoA protein - Burkholderia cep
pir2:B44390	+ 112.00	138.17	1.65	802	i hypothetical protein 33L - Moll
					i protein-tyrosine-phosphatase (B

pir2:C75580	+ 111.50	144.86	1.49	376	i adenine deaminase-related
pir2:JC4176	+ 111.00	136.61	1.98	817	i pyruvate,water dikinase (EC
pir2:T45134	+ 108.50	140.15	1.94	529	i hypothetical protein [import
pir2:T34434	+ 108.50	123.36	3.96	2232	i hypothetical protein K06A9.
pir2:T50711	+ 108.00	135.99	3.08	569	i urease (EC 3.5.1.5) alpha ch
pir2:S25370	- 107.00	126.51	4.52	1306	i MSB2 protein - yeast (Sacch
pir2:S48478	- 107.00	126.06	4.57	1367	i glucan 1,4-alpha-glucosidas
pir2:B46629	- 106.50	135.08	3.90	505	i mucin 6, gastric (3-repeat c
pir2:A39344	- 106.50	132.92	4.12	630	i tumor-associated mucin (MUC1
pir2:H70108	+ 106.50	130.93	4.33	773	i hypothetical protein BB0072
pir2:T28969	+ 106.50	121.75	5.89	1847	i hypothetical protein T23H2.
pir2:T34433	- 105.50	126.73	5.56	1032	i hypothetical protein K06A9.
pir1:PWNTB	+ 105.00	133.14	5.07	498	i H+-transporting two-sector A
pir1:PWNTB	+ 105.00	133.14	5.07	498	i H+-transporting two-sector A
pir1:PWNTB	+ 105.00	128.27	5.74	821	i pyruvate,water dikinase (EC
pir1:PWNTB	+ 105.00	126.45	6.01	989	i insulin II gene enhancer-bin
pir1:PWNTB	+ 105.00	113.94	8.29	3570	i mucin MUC5B, tracheobronchi
pir1:PWNTB	+ 104.50	135.53	5.11	363	i branched-chain amino acid am
pir1:PWNTB	+ 104.50	133.52	8.99	3473	i genome polyprotein - rice t
pir1:PWNTB	+ 104.00	132.60	5.92	457	i UDP-N-acetylmuramoylalanyl-D
pir1:PWNTB	- 103.50	126.55	7.42	792	i vesicular transport protein
pir1:PWNTB	- 103.50	125.22	7.68	907	i membrane antigen gp350 - hum
pir1:PWNTB	+ 103.50	116.41	9.63	2241	i hypothetical protein UL48 -
pir1:PWNTB	+ 103.00	114.25	10.17	2796	i fatty-acid synthase (EC 2.3.
pir1:PWNTB	+ 103.00	130.47	7.20	493	i aldehyde dehydrogenase dh
pir1:PWNTB	+ 103.00	129.25	7.43	559	i aldehyde dehydrogenase dh
pir1:PWNTB	+ 102.50	118.29	10.57	1603	i hypothetical protein Vng0553
pir1:PWNTB	- 102.50	133.40	7.70	317	i gag/pol polyprotein - avian
pir1:PWNTB	+ 102.00	133.06	7.77	328	i gag/pol polyprotein - pig
pir1:PWNTB	+ 102.00	128.99	8.62	498	i salivary glue protein sgs-3
pir1:PWNTB	+ 102.00	128.99	8.62	498	i H+-transporting two-sector A
pir1:PWNTB	+ 102.00	125.02	9.55	749	i H+-transporting two-sector A
pir1:PWNTB	- 102.00	124.31	9.72	805	i hypothetical protein AP0340
pir1:PWNTB	+ 102.00	123.25	9.99	898	i hypothetical protein F53F10.
pir1:PWNTB	+ 102.00	128.11	9.47	508	i H+/K+-exchanging ATPase (EC
pir1:PWNTB	+ 101.50	127.75	9.56	527	i histidine ammonia-lyase (EC
pir1:PWNTB	- 101.50	96.30	21.41	13288	i translation initiation facto
pir1:PWNTB	+ 101.50	131.25	9.38	343	i mucin, submaxillary - pig
pir1:PWNTB	+ 101.00	121.96	11.02	654	i acidic ribosomal protein P0
pir1:PWNTB	+ 101.00	121.96	11.02	654	i hypothetical protein CG2730
pir1:PWNTB	+ 101.00	120.91	11.98	914	i cation-transporting ATPase.
pir1:PWNTB	- 101.00	120.91	12.22	990	i probable proline dehydrogena
pir1:PWNTB	+ 100.50	130.30	10.31	352	i probable membrane protein Y
pir1:PWNTB	+ 100.50	122.97	10.40	364	i probable protein of beta-pro
pir1:PWNTB	+ 100.50	122.56	12.58	779	i phospho-2-dehydro-3-deoxyhep
pir1:PWNTB	+ 100.50	117.99	14.14	1245	i hypothetical protein PA3909
pir1:PWNTB	+ 100.50	117.18	14.44	1245	i hypothetical protein CG0739
pir1:PWNTB	+ 100.50	115.69	15.00	1576	i hypothetical protein W09010
pir1:PWNTB	+ 100.00	129.18	11.40	368	i pol protein - yeast (Candid
pir1:PWNTB	+ 100.00	125.07	12.66	561	i hypothetical protein PH0483
pir1:PWNTB	+ 100.00	124.84	12.74	574	i hypothetical protein SC9A10.
pir1:PWNTB	+ 100.00	119.25	14.70	1019	i ABC-type transport protein s
pir1:PWNTB	- 100.00	114.80	16.48	1609	i protein R52.2 [imported] -
pir1:PWNTB	- 99.50	134.88	10.57	191	i probable membrane protein Y
pir1:PWNTB	+ 99.50	129.91	12.01	318	i hypothetical protein - Strept
pir1:PWNTB	+ 99.50	126.66	13.05	444	i membrane fusion protein (ntr
pir1:PWNTB	+ 99.50	125.88	13.31	481	i probable IMP dehydrogenase
pir1:PWNTB	+ 99.50	125.02	13.61	525	i H+-transporting two-sector A
pir1:PWNTB	+ 99.50	125.02	13.61	525	i hypothetical 57.4K protein (

seq_name: pir2:A81919

seq_documentation_block:

probable sodium-translocating NADH dehydrogenase (ubiquinone) (EC 1.6.5.-) chain A NM
C:Species: Neisseria meningitidis
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 16-Feb-2001
C:Accession: A81919
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mo
; Holroyd, S.; Jags, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandre
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491
A:Reference number: A81775; MUID:20222556

A:Accession: A81919
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-447 <PAR>
A:Cross-references: GB:AL162754; GB:AL157959; NID:g7379424; PIDN:CAB84035.1; PID:g737947
A:Experimental source: serogroup A, strain 22491
C:Genetics:
A:Gene: nqrA; NMA0752
C:Superfamily: Vibrio alginolyticus sodium-translocating NADH dehydrogenase (ubiquinone)
C:Keywords: NAD; oxidoreductase

alignment_scores:
Quality: 2222.00 Length: 447
Ratio: 4.982 Gaps: 0
Percent Similarity: 99.776 Percent Identity: 96.197

alignment_block:
US-09-303-518d-131 x A81919 ..

Align seg 1/1 to: A81919 from: 1 to: 447

```
1 ATGATTAAATCAAAAAGCTTAATCTGCCATCGCGGCGAGACCGGA 50
1 MetIleIysIleIysIysGlyLeuAsnLeuProIleAlaGlyArgProG1 17
51 GCAAGTCATTATGACGGCCCGGCCCAATTACCGAAGTCGGTTCGTCGGC 100
17 uGlnValIleTyrAspGlyProValIleThrGluValAlaLeuLeuGlyG 34
101 AAGAATATGTCGGCATCGCCCTCGCATGAAATCAAGGAAGGTGAAGCC 150
34 luGluTyrAlaGlyMetArgProSerMetIysValIysGluGlyAspAla 50
151 GTCAAAAAGGCCCAAGTCGTGTTTGAAGACAAAGAAATCCGGCGCTAGT 200
51 ValIysIysGlyGlnValLeuPheGluAspLysLysAsnProGlyValVa 67
201 ATTACTGCGCGCTTCAGGCAAAATCGCGCTATTACCGGCGCGAA 250
67 lPheThrAlaProValSerGlyIysIleAlaIleHisArgGlyGluL 84
251 AGCGCGTACTTCAGTCAGTCGTGATTGCCGTGGAAGCAACACGAATC 300
84 ysArgValLeuGlnSerValValIleAlaValGluGlyAsnAspGluIle 100
301 GAGTTCGAAGCTACGTACCTGAGCGCTGCGCAAAATTTGACGCGAAA 350
101 GluPheGluArgTyrAlaProGluAlaLeuAlaAsnLeuSerGlyGluG 117
351 AGTCCGCGCAACCTGATTCAATCAGGCTTATGACTCGCTTCGCAACC 400
117 uValArgArgAsnLeuIleGlnSerGlyLeuTyrThrAlaLeuArgThrA 134
401 GTCGTTTACGAAATCCCTGCGGTAGATGCGGAGCGCTTCGCACTTC 450
134 rgProPheSerLysIleProAlaValAspAlaGluProPheAlaIlePhe 150
451 GTCAATGCGATGGACACCAATCCGCTGGCTGCGGACCCCTACGGTCAAT 500
151 ValAsnAlaMetAspThrAsnProLeuAlaAlaAspProValValValI 167
501 CAAAGAAGCCCGCAAGACTTCAACACGCGGCTGTTGTTGTTGAGCGCC 550
167 eLysGluAlaAlaGluAspPheArgArgGlyLeuLeuValLeuSerArgL 184
551 TGACCGAACGTAATAATCCATGTTGTGAACGACGAGCGGAGAGCTGCC 600
184 euThrGluArgLysIleHisValCysLysAlaAlaGlyAlaAspValPro 200
601 TCTGAAATGCTGCAATATCGAATACACATGAATTTGGCGCGCCGATCC 650
201 SerGluAsnAlaAlaAsnIleGluThrHisGluPheGlyProHisPr 217
```

```
651 TGCCGGCTTGAGTGGCAGCACATTCATTTTCATCGAGCCAGTCGCGCGA 700
217 oAlaGlyLeuSerGlyThrHisIleHisPheIleGluProValGlyAla 234
701 ATAAACCGTGTGGACCATCAATATATCAAGACGTGATTCATATCGACGT 750
234 snLysThrValTyrThrIleAsnTyrGlnAspValIleAlaIleGlyArg 250
751 TTGTTCTATACAGCCCTCTGATATACCGACGCGGTGTTGCCCTTGGCGG 800
251 LeuPheAlaThrGlyArgLeuAsnThrGluArgValIleAlaLeuGlyG1 267
801 CCTGCAAGTCAACAAACCGCGCTCTTTCGCTACCGTTCCTGGTGGTGAAG 850
267 ySerGlnValAsnLysProArgLeuLeuArgThrValLeuGlyAlaLys 284
851 TGCTCAACTTACCGCGCGCAATTTGTTGACCGGGACAAACCGGTGATT 900
284 alSerGlnIleThrAlaGlyGluLeuValAspAlaAspAsnArgValIle 300
901 TCCGTTTCGGTATTGAACGGTGGATTGCACAGGCGCGCATGATTATTT 950
301 SerGlySerValLeuAsnGlyAlaIleThrGlnGlyAlaHisAspTyrLe 317
951 GGGACGCTTACCAATCAGATTTCCGTTTATCGAAGAGCGCGCAGCAAG 1000
317 uGlyArgTyrHisAsnGlnIleSerValIleGluGluGlyArgSerLys 334
1001 AGCTGTTGCGGTGGTTCGCGCGCGGACGCAAAATATCTCATCAGCGC 1050
334 luLeuPheGlyTyrValAlaProGlnProAspLysTyrSerIleThrArg 350
1051 ACCACTCGCGCCATTTCTTAAAAAACAACCTCTCAAGTTCACGACAGC 1100
351 ThrThrLeuGlyHisPheLeuLysAsnLysLeuPheLysPheThrAl 367
1101 CGTCAACGCGCGGACCGCGCATGGTACCGATCGGCATTCATGACGCG 1150
367 aValAsnGlyGlyAspArgAlaMetValProIleGlyThrTyrGluArg 384
1151 TAATCCGCTTGACATCTCTGCTACCTGCTTTCGCGGATTTAATGTC 1200
384 alMetProLeuAspIleLeuProThrLeuLeuLeuArgAspLeuIleVal 400
1201 GCGCATACGACGCGCGCAGCGCTTGGTGGCTTGGATTTGGACGAGA 1250
401 GlyAspThrAspSerAlaGlnAlaLeuGlyCysLeuGluLeuAspGluG 417
1251 AGACCTCGCTTGTGACGCTGCTGCTGCGCGGCAAAATACGAATACGCG 1300
417 uAspLeuAlaLeuCysSerPheValCysProGlyLysTyrGluTyrGly 434
1301 CGCTGTTGCGCAAGTGTGGAACCATTTGAGAGGAAGGC 1341
434 roleuLeuArgLysValLeuGluThrIleGluLysGluGly 447
```

seq_name: pir2:D81185

seq_documentation_block:
sodium-translocating NADH dehydrogenase (ubiquinone) (EC 1.6.5.-) alpha chain NMB0569
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 02-Mar-2001
C:Accession: D81185
R:Tetelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, R.; Tetelin, H.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B. H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizzz, M. Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: A81000; MUID:20175755
A:Accession: D81185
A:Status: preliminary

A:Molecule type: DNA
A:Residues: 1-447 <TET>
A:Cross-references: GB:AE002412; GB:AE002098; NID:g7225783; PIDN:AAF40997.1; PID:g722579
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB0569
C:Superfamily: Vibrio alginolyticus sodium-translocating NADH dehydrogenase (ubiquinone)
C:Keywords: oxidoreductase

alignment_scores:
Quality: 2221.00 Length: 447
Ratio: 4.980 Gaps: 0
Percent Similarity: 99.776 Percent Identity: 96.197
alignment_block:
us-09-303-518d-131 x D81185 ..
Align seg 1/1 to: D81185 from: 1 to: 447
1 ATGATTAAATCAAAAAGGCTTAATCTGCCCATCGCGGAGACCGGA 50
1 MetIleLysIleLysLysGlyLeuAsnLeuProIleAlaGlyArgProI 17
51 GCAAGTCATTATGACGGCCCGCCATACCGAAGTCGCGTTGCGG 100
17 uGlnAlaValTyrAspGlyProAlaIleThrGluValAlaLeuLeuGlyG 34
101 AAGAATATCTGCGCATGCGCCCTCGATGAAATCAAGGAAGTGAAGCC 150
34 luGluTyrAlaGlyMetArgProSerMetLysValLysGluGlyAspAla 50
151 GTCAAAAAGGCGAAGTCTGTTTGAAGACAAAAAGAAATCCGGCGTAGT 200
51 ValLysLysGlyGlnValLeuPheGluAspLysLysAsnProGlyValVa 67
201 ATTTACTGCGCGCTTCAGGCAAAATCCCGCTATTACCGTGGCGAAA 250
67 lPheThrAlaProAlaSerGlyLysIleAlaAlaIleHisArgGlyGluL 84
251 AGCGCGTACTTCAGTCAGTCTGATGTCGCTTGAAGCAAGCAAGCAATC 300
84 ysArgValLeuGlnSerValValIleAlaValGluGlyAsnAspGluIle 100
301 GAGTTCGAACGTACGTACCTGAAGCGCTGCGCAAAATGAGCAGCAAAA 350
101 GluPheGluArgTyrAlaProGluAlaLeuAlaAsnLeuSerGlyGluG 117
351 AGTGGCGCGCAACCTGATTCAATCAGGCTTATGGACTCGCGTTGCGACC 400
117 uValArgArgAsnLeuIleGlnSerGlyLeuTrpThrAlaLeuArgThrA 134
401 GTCGCTTCAGCAAAATCCCTGCTAGTACGCGCGCTTCGCATCTTC 450
134 rgProPheSerLysIleProAlaValAspAlaGluProPheAlaIlePhe 150
451 GTCATGCGATGGACCAATCCGCTGCTGCGACCCCTAGCGTCATCAT 500
151 ValAsnAlaMetAspThrAsnProLeuAlaAlaAspProThrValIleI 167
501 CAAAGAACGCCCGCAAGACTTCAACCGCGCTGTTGTTGTTGAGCGGCC 550
167 eLysGluAlaAlaGluAspPheLysArgGlyLeuLeuValLeuSerArgL 184
551 TGACCGAACGTAAATCCATGTGTGTAAGCAGCAGCGCAGAGCGTCCG 600
184 euThrGluArgLysIleHisValCysLysAlaAlaGlyAlaAspValPro 200
601 TCTGAAATGCTGCAATATCGAACAACATCAATTTGCGCGCCCGCATCC 650
201 SerGluAsnAlaAlaAsnIleGluThrHisGluPheGlyGlyProHisPr 217
651 TGCGCGCTTGAGTGGCAGCAGACATTCATTTCATCGAGCCAGTCGCGCGA 700

|||||
217 oLaGlyLeuSerGlyThrHisIleHisPheIleGluProValGlyAlaA 234
701 ATAAACCCGTGGACCATCAATTAACAGCGTGAATTCCTATCGGAGGT 750
|||||
234 snLysThrValTrpThrIleAsnTyrGlnAspValIleThrIleGlyArg 250
751 TTGTTCTGAACAGCGCTCTGAATACCGAGCGCTGGTTCCTTGGCGGG 800
251 LeuPheAlaThrGlyArgLeuAsnThrGluArgValIleAlaLeuGlyG 267
801 CCTGCAAGTCACAACACCGCCCTCTTCGCTACCGCTTTGGTGGCGAAG 850
267 ySerGlnValAsnLysProArgLeuLeuArgThrValLeuGlyAlaLysV 284
851 TGTCTCAACTTACCGCGCGCAATTTGTTGACGGGACCAACCGCTGATT 900
284 alSerGlnIleThrAlaGlyGluLeuValAspThrAspAsnArgValIle 300
901 TCCGTTCCGTTATTGAACGGTGCATTGCACAGCGCGCATGATTATTT 950
301 SerGlySerValLeuAsnGlyAlaIleThrGlnGlyAlaHisAspTyrLe 317
951 GGGAGCTTACCACATCAGATTTCCTGATTCGAAAGAGCGCGCAGCAAG 1000
317 uGlyArgTyrHisAsnGlnIleSerValIleGluGlyArgSerLysG 334
1001 AGCTGTTGCGTGGTTCGCGCGAGCGGACAAATACTCCATCACGCGC 1050
334 luLeuPheGlyTyrValAlaProGlnProAspLysTyrSerIleThrArg 350
1051 ACCACTCTCGCCCATTTCTTAAACAACTCTTCAAGTTCACGACAGC 1100
351 ThrThrLeuGlyHisPheLeuLysAsnLysLeuPheLysPheAsnThrAl 367
1101 CGTCAACGGCGCGACCGCCATGTTACGATCGCGCCTTATGACGCGG 1150
367 aValAsnGlyGlyAspArgAlaMetValProIleGlyThrTyrGluArgV 384
1151 TAATGCGTGGACATCCTGCTACCTGCTTTGCGCGATTTAATCGTC 1200
384 alMetProLeuAspIleLeuProThrLeuLeuLeuArgAspLeuIleVal 400
1201 GCGGATACCGACAGCGCGAGCTTTGGTTCCTTGAATTCGACGAAAG 1250
401 GlyAspThrAspSerAlaGlnAlaLeuGlyCysLeuGluLeuAspGluG 417
1251 AGACCTCGCTTTGTCAGCTTCGCTGCGCGGCAATACGAATACGCGC 1300
417 uAspLeuAlaLeuCysSerPheValCysProGlyLysTyrGluTyrGlyP 434
1301 CGCTGTTGCCCAAGTGTGGAACCATTTGAGAAGGAAGGC 1341
434 roLeuLeuArgLysValLeuGluThrIleGluLysGluGly 447
seq_name: pir2:164002

seq_documentation_block:
Sodium-translocating NADH dehydrogenase (ubiquinone) (EC 1.6.5.-) alpha chain HI0164
C:Species: Haemophilus influenzae
C:Date: 18-Aug-1995 #sequence_revision 02-Sep-2000 #text_change 02-Mar-2001
C:Accession: I64002; A64003
R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman
D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350630
A:Accession: I64002
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-447 <TIGR>

A:Cross-references: GB:L42023; TIGR:HI0164; GB:U32702; NID:gl573118; PIDN:AAC21836.1; PID:gl573122
 A:Note: the sequence is revised in GenBank entry U32702, PID:gl573122
 C:Superfamily: Vibrio alginolyticus sodium-translocating NADH dehydrogenase (ubiquinone)
 C:Keywords: oxidoreductase

alignment_scores:
 Quality: 1646.00 Length: 448
 Ratio: 4.146 Gaps: 2
 Percent Similarity: 88.616 Percent Identity: 70.089

alignment_block:
 US-09-303-518d-131 x I64002 ..

Align seg 1/1 to: I64002 from: 1 to: 447

```

1 ATGATTAATAATCAAAAGAGTCTAAATCTGCCATCGCGGCGAGCCGGA 50
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
1 MetileThrIleLysGlyLeuAspLeuProIleAlaGlyLysProAl 17
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
51 GCAAGTCATTATGACGGCGCGCCATTACCGAAGTCGCGTTCGTTGGCG 100
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
17 aGlnValIleHisSerGlyAsnAlaValAsnGlnValAlaIleLeuGly 34
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
101 AGAATATGTCGCATCGCGCCCTCGATGAAATCAAGGAGGTGAAGCC 150
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
34 luGluTyrValGlyMetArgProSerMetLysValArgGluGlyAspVal 50
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
151 GTCAAAAAGCGCAAGTCGTTGTAAGACAAAAGAAATCCGGCGGTAGT 200
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
51 ValLysLysGlyGlnValLeuPheGluAspLysLysAsnProGlyValI 67
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
201 ATTACTGCGCGGCTTCAGGCAAAATCGCGCTATTCACCGTGGCGAAA 250
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
67 ePheThrAlaProIleSerGlyThrIleThrAlaIleAsnArgGlyGlu 84
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
251 AGCGGTACTTCAGTCAGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 300
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
84 ysArgValLeuGlnSerValValIleAsnValGluGlyAspGluLysIle 100
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
301 GAGTTCGAACGCTACGTACCTGAGCGCTGCGCAAAATGAGCGCGNAAA 350
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
101 ThrPheAlaLysTyrSerThrGluGlnLeuAsnThrLeuSerSerGluG 117
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
351 AGTCGCGCGCAACCTGATCAATCAGCTGATGACTGCGCTTCGCGACCC 400
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
117 nValLysGlnAsnLeuIleGluSerGlyLeuThrPheAlaLeuArgThrA 134
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
401 GTCGCTTCAGCAAAATCCCTGCGCTAGATGCCGCGCTTCGCGCATCTC 450
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
134 rgProPheSerLysValProSerIleGluSerGluAlaSerSerIlePhe 150
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
451 GTCATGCGGATGGACCAATCCGCTGCTGCCACCTACGCTACGTCATCAT 500
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
151 ValAsnAlaMetAspThrAsnProLeuAlaAlaAspProSerValValLe 167
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
501 CAAGAAGCCCGGAGACTTCAACAGCGGCTGTGTTGTTATGAGCGGCC 550
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
167 uLysGluTyrSerGlnAspPheThrAsnGlyLeuThrValLeuSerArgL 184
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
551 TGACCGAACGTA...ATCATGTGTGTAAGCAGCAGCGCGCAGACGTG 597
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
184 euPheProSerLysProLeuHisLeuCysLysAlaGlyAspSerAsnIle 200
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
598 CCGTCTGAAATGCTGCCAATATCGAACAACATGAATTTGCGCGCCGCA 647
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
201 ProThrAlaAspLeuGluAsnLeuGlnIleHisaspPheThrGlyValH 217
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
648 TCCTCGCGCTTGTAGTCAGCAGCATATTCATTCATCGAGCAGTCGCGG 697
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
217 sProAlaGlyLeuValGlyThrHisIleHisPheIleAspProValGlyI 234
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
698 CGAATAAAACCGTGTGGACCATCAATATCAAGACGTGATTGCTATCGGA 747

```

```

234 leGlnLysThrValThrPheIleAsnTyrGlnAspValIleAlaValGly 250
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
748 CGTTTGTTCGTACAGCCGCTCGAATPACCGAGCGGCTGTTGCCCTGGG 797
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
251 LysLeuPheThrThrGlyGluLeuTyrSerGluArgValIleSerLeuAl 267
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
798 CGGCTGCAAGTCAACAAACCGCGCTTCGTCACCGTTTGGTGGCGGA 847
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
267 aglyProGlnValLysGluProArgLeuValArgThrThrIleGlyAlaA 284
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
848 AGTGTCTCAACTTACCGCGCGAATTTGTTGACGCGGACAAACCGCGTG 897
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
284 snLeuSerGlnLeuThrGlnAsnGluLeuSerAlaGlyLysAsnArgVal 300
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
898 ATTTCCGCTCGGTATTCGACGGTGCATTCACAGCGCGCGCATGATTA 947
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
301 IleSerGlySerValLeuCysGlyGlnIleAlaLysAspSerHisAspTy 317
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
948 TTGCGGCGCTACCAATCAGATTTCGTTATCGAAGAGCGCGCGACGA 997
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
317 rLeuGlyArgTyrAlaLeuGlnValSerValIleAlaGluGlyAsnGlu 334
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
998 AAGAGCTGTCGCTGCGTTCGCGCGAGCGGACGACAAATACTCCATCAG 1047
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
334 ysGluPheGlyTrpIleMetProGlnAlaAsnLysTyrSerValThr 350
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
1048 CGCACCTCTCGGCAATTCCTTAAACAACTCTTCAAGTTTCAGGAC 1097
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
351 ArgThrValLeuGlyHisPheSerLys...LysLeuPheAsnPheThrTh 366
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
1098 AGCCGCTCAACGCGCGCGCATGTCACCGATCGGACGCTTATGAGC 1147
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
366 rSerGluAsnGlyGlyGluArgAlaMetValProIleGlySerTyrGluA 383
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
1148 GCGTAATGCGCTGACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1197
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
383 rgValMetProLeuAspIleLeuProThrLeuLeuLeuArgAspLeuIle 399
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
1198 GTCGCGCATACCGACGCGCGCGCTTGGTTCGTTGGAATGGAGCA 1247
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
400 ValGlyAspThrAspGlyAlaGlnGluLeuGlyCysLeuGluLeuAspG 416
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
1248 AGAAGACCTCGCTTGTGCGAGCTTGTGCTGCGCGGCAATACGAAATAC 1297
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
416 uGluAspLeuAlaLeuCysSerPheValCysProGlyLysTyrGluTyrG 433
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
1298 GCGCGCTGTCGCAAGTCTGGAACCATTCGAGAGGAAGGC 1341
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
433 lySerIleLeuArgGlnValLeuAspLysIleGluLysGluGly 447
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

```

seq_name: pir2:AG0393

seq_documentation_block:
 NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain A [imported] - Yersinia pestis (st
 C:Species: Yersinia pestis
 C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 27-Nov-2001
 C:Accession: AG0393
 R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.
 geno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.
 il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrel
 Nature 413, 523-527, 2001
 A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
 A:Reference number: AB0001; MUID:21470413; PMID:11586360
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-447 <KUP>
 A:Cross-references: GB:AL590842; PIDN:CAC92475.1; PID:gl5981176; GSPDB:GN00175
 C:Genetics:
 C:Gene: nqrA
 C:Superfamily: Vibrio alginolyticus sodium-translocating NADH dehydrogenase (ubiquino
 C:Keywords: oxidoreductase

```

|||||... |||||... |||||... |||||... |||||... |||||... |||||... |||||...
251 LeuPheThrArgGlyGluLeuCysThrAspArgIleValAlaLeuAlaGl 267

      CCTCGAAGTCAACAACCCGGCCCTTCGTACCGTTTTGGTGCGGAAG 850
      |.....|.....|.....|.....|.....|.....|.....|.....|
267 yProGlnValAsnGlnProIleLeuLeuArgThrArgLeuGlyAlaSerL 284

      TGTCTCAACTACC GCCGCCGAATTGTTGACGCCGACAACCCGCGTATT 900
      eSerGluLeuThrAlaGlyLysLeuLysGluGlyAspAsnArgIleIle 300

      TC CGGTTTCGATTGAACGGTGCATTCACAAAGCCGCCGATGATTATT 950
      |||||... |||||... |||||... |||||... |||||... |||||... |||||...
301 SerGlySerValLeuSerGlyThrAlaPheSerAlaThrHisGlyTy rLe 317

      GGACGCTACCAAAATCAGATTTCCTGTATCGAAGAGGCCGACGACAAAG 1000
      |||||... |||||... |||||... |||||... |||||... |||||... |||||...
317 uGlyArgPheHisGlnValSerValIleArgGluGlyArgGluLys 334

      AGCTGTTGGCTCGGTGGCCGCGCACCGCGACAATACTCCATCAGCGCG 1050
      lLeuPheGlyTrpValMetProGlyArgAspLysTyrSerIleThrArg 350

      ACCACTCTCGGCATTTCCTAAAAACAACACTCTTCAAGTTCACGACGC 1100
      |||||... |||||... |||||... |||||... |||||... |||||... |||||...
351 ThrThrLeuGlyHisPhePheYsrGlyLeuPheAlaPheSerThrAs 367

      CGTCAACGGCGCGCACCGCGCATGTACCGATCGGCACCTTATGACGGCG 1150
      :...:~...:~...:~...:~...:~...:~...:~...:~...:~...:~...:~...:~...
367 pMethHisGlyGlyLuargAlametValProIleGlyAsnTyrGluArgv 384

      TTAATCGCGTTGGACAPCTCGCTACCTTCTGCTTTGCGCGATTAAATCGTC 1200
      aIMetProLeuAspIleLeuAlaThrHisLeuLeuArgAspLeuLeuAla 400

      GCGGATACCGACAGCGCGACGCTTGGGTTCTTGGAAATTGACGAGAAGA 1250
      401 GlyAspThrAspSerAlaGlnAlaLeuGlyCysLeuGluLeuAspGluGI 417

      AGACCTCGCTTTGTGCAGCTTCGCTCGCCCGGCAATAACGAATACGGCC 1300
      |||||... |||||... |||||... |||||... |||||... |||||... |||||...
417 uAspLeuAlaLeuCysThrPheValCysProGlyLysTyrGluTy rGlyP 434

      CGCTCTTCGGCAAGTGCTGGAAACCATTTGACAGGAAGGC 1341
      ||:::||||| ~~~~~~ |||||... |||||... |||||... |||||...
434 roValLeuArgspIleuThrGlnIleGluGlnGly 447

```

```

seq_name: pir2:S51015
seq_documentation_block:
sodium-translocating NADH dehydrogenase (ubiquinone) (EC 1.6.5.-) alpha chain nqra [
N;Alternate names: nqra protein
C;Species: Vibrio alginolyticus
C;date: 10-Apr-1996 #sequence_revision 19-Apr-1996 #text_change 16-Feb-2001
C;Accession: S51015; S66365
R;Beattie, P.; Tan, K.; Bourne, R.M.; Leach, D.; Rich, P.R.; Ward, F.B.
FEBS Lett. 356, 333-338, 1994
A;Title: Cloning and sequencing of four structural genes for the Na(+)-translocating
A;Reference number: S51013; MUID:95104445
A;Accession: S51015
A;Molecule type: DNA
A;Residues: 1-446 <BEA>
A;Cross-references: EMBL:Z37111; NID:g663269; PIDN:CAA85476.1; PID:g663270
A;Accession: S66365
A;Molecule type: protein
A;Residues: 1-10 <BEW>
C;Genetics:
A;Gene: nqra
C;Superfamily: Vibrio alginolyticus sodium-translocating NADH dehydrogenase (ubiquinone)
C;Keywords: FMN; oxidoreductase
alignment_scores:

```

Quality: 1488.50 Length: 447
Ratio: 3.856 Gaps: 1
Percent Similarity: 86.353 Percent identity: 63.758

alignment_block:
US-09-303-518D-131 x S51015 ..
Align seg 1/1 to: S51015 from: 1 to: 446

1 ATGTAATAAATAAAGGCTTAATCTGCCATCGCGGGCAGACCGA 50
1 MetilleThrIleLysLysGlyLeuAspLeuProIleAlaGlyThrProSe 17
51 GCAAGTCATTATGACGGCCCGCCATACCGAAGTCGGTGGCGG 100
17 rGlnValIleAsnAspGlyLysThrIleLysLysValAlaLeuGlyG 34
101 AAGAATATGTCGCGATCGCCCTCGATCAAAATCAAGGAAGTGAAGCC 150
34 luGluThrValGlyMetArgProThrMetHisValArgValGlyAspGlu 50
151 GTCAAAAAGGCAAGTGTGTGTAAGACAAAGAAAGATCCGGCGTAGT 200
51 ValLysLysAlaGlnValLeuPheGluAspLysLysAsnProGlyVally 67
201 ATTACTGCGCGCTTCAGGCAAAATCGCCGTATTACCGTGGCGAAA 250
67 sPheThrAlaProAlaAlaGlyLysValIleGluValAsnArgGlyAlaL 84
251 AGCGGTACTTTCAGTCAGTGTGATGCGGTGAAGCAACAGCAAGATC 300
84 ysargValLeuGlnSerValValIleGluValAlaGlyGluGlnVal 100
301 GAGTTCGACGCTACGTACCTGAGCGCTGCGCAAAATTCAGCAGCAAAA 350
101 ThrPheAspLysPheGluAlaAlaGlnLeuSerGlyLeuAspArgGluVa 117
351 AGTGGCGCGCAACTGATTCATCAATCAGCTTATGACTGCGCTTCGCACCC 400
117 lIleLysThrGlnLeuValAspSerGlyLeuThrAlaLeuArgThra 134
401 GTCCGTTACAGAAATCCCTGCGCTAGATCCGAGCGCTTCGCATCTTC 450
134 rgProPheSerLysValProAlaIleGluSerThrLysAlaIlePhe 150
451 GTCATCGGATGACACCAATCGCTGCGTGCAGCCCTACGTCATCAT 500
151 ValThrAlaMetAspThrAsnProLeuAlaAlaLysProGluLeuIle 167
501 CAAGAAGACCGCCGCAAGACTTCAACCGCGCTGTGTTGTTAGCGCGCC 550
167 eAsnGluGlnGluAlaPheIleAlaGlyLeuAspIleLeuSerAlaL 184
551 TGACCGAAGCTAAATCCATGTGTGTAACAGCAGCAGCGCGCGCGCG 600
184 eutThrGluGlyLysValThrValCysLys...SerGlyThrSerLeuPro 199
601 TCTGAATATGTCGAATATGCAATATGCAATATGCAATATGCAATATG 650
200 ArgSerGlnSerAsnValGluGluHisValPheAspGlyProHisPr 216
651 TGCGCGGTGTAGTGGCAGCATATTCATTCATCGACCGCAGCGCGCGA 700
216 oAlaGlyLeuAlaGlyThrHisMetHisPheLeuThrProValAsnAlaG 233
701 ATAAACCGGTGGACCATCAATATTCAGACGCTGATTCGTCGACGCT 750
233 luAsnValAlaTrpSerIleAsnThrGlnAspValIleAlaPheGlyLys 249
751 TTGTTCTGACAGCGCTCAATACCGCGCTGTTGCTTGGCGGCGG 800
250 LeuPheLeuThrGlyGluLeuThrThrAspArgValValSerLeuAlaG 266

801 CCTGCAAGTCAACAAACCGCGCTCTTGGTACCGTTTGGGTGCGAAG 850
266 yProValValAsnAsnProArgLeuValArgThrValIleGlyAlaSerL 283
851 TGCTCTCAACTTACCGCGCGAATTTGGTTCACCGCGGACACCGCGTATT 900
283 euAspAspLeuThrAspAsnGluLeuMetProGlyGluValArgValIle 299
901 TCCGTTCCGTTATGAACGTCGATTGTCACAAAGCGCGCATGATTATT 950
300 SerGlySerValLeuThrGlyThrHisAlaThrGlyProHisAlaThrLe 316
951 GGGACGCTACCAATCAGATTTCCGTTATCGAAGAGCGCGCAGCAAG 1000
316 uGlyArgThrHisGlnGlnValSerValLeuArgGluGlyArgGluLys 333
1001 AGCTGTTCCGCTGGTTCGCGCGCAGCGCGCAAAATCTCCATCAGCGC 1050
333 luLeuPheGlyTrpAlaMetProGlyLysAsnLysPheSerValThrArg 349
1051 ACCACTCTCGGCCCATTTCTTAAAAAACAACACTCTCAAGTTCAGCAGC 1100
350 SerPheLeuGlyHisValPheLysGlyGlnLeuPheAsnMetThrThr 366
1101 CGTCAACGGCGCGACCGCGCCATCGTACCGTCCGCGCCTTATCAGCGC 1150
366 rThrAsnGlySerAspArgSerMetValProIleGlyAsnThrGluArg 383
1151 TAATGCGCTGTCGATCTGCTACCTTGTGTCGCGCATTTAATCGTC 1200
383 alMetProLeuAspMetGluProThrLeuLeuLeuAspLeuCysAla 399
1201 GCGGATACCGACGCGCGCGCTTGGTGTGCTTGAATTTGGACGAGA 1250
400 GlyAspThrAspSerAlaGlnAlaLeuGlyAlaLeuGluLeuAspGlu 416
1251 AGACCTCTGCTTTCGCTGCTGCTGCGCGCAAAATACGAATACGCGC 1300
416 uAspLeuAlaLeuCysThrPheValCysProGlyLysThrGluThrGly 433
1301 CGCTGTTGCGCAAGTGTGGAACCATTTGAGAAGGAGGC 1341
433 hrLeuLeuArgGluCysLeuAspThrIleGluLysGluGly 446

seq_name: pir2:G82094
seq_documentation_block:
sodium-translocating NADH dehydrogenase (ubiquinone) (EC 1.6.5.-) nqrA VC2295 [import
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 16-Feb-2001
C:Accession: G82094
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers
L, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833
A:Accession: G82094
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-462 <HEI>
A:Cross-references: GB:AF004300; GB:AE003852; NID:g9656850; PIDN:AAF95439.1; GSPDB:GN
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC2295
A:Map position: 1
C:Superfamily: Vibrio alginolyticus sodium-translocating NADH dehydrogenase (ubiquino
C:Keywords: Oxidoreductase

alignment_scores:
Quality: 1476.50 Length: 447
Ratio: 3.845 Gaps: 1

Percent Similarity: 85.906 Percent Identity: 62.416

alignment_block:

US-09-303-518D-131 x G82094

Align seg 1/1 to: G82094 from: 1 to: 462

```

1  ATGATTAAATCAAAAAGGTCTAAATCTGCCCATCGCGGCGGACACCGGA 50
   |||||
17  MetIleThrIleLysLysGlyLeuAspLeuProIleAlaGlyThrProSe 33
   |||||
51  GCAGTCAATTTATGACGGCCGGCCCATACCGAAGTCGCTTGGTGGCG 100
   |||||
33  rGlnValIleSerAspGlyLysAlaIleLysLysValAlaLeuLeuGly 50
   |||||
101  AAGAATATCTGGCATCGCCCTCGATGAAATCAAGAGGTGAAGCC 150
   |||||
50  luGluTyValGlyMetArgProThrMetHisValArgValGlyAspGlu 66
   |||||
151  GTCAAAAAGGCCAAGTCTGTTGAAGACAAAGAAATCCGGCGGTAGT 200
   |||||
67  ValLysLysAlaGlnIleLeuPheGluAspLysLysAsnProGlyValLy 83
   |||||
201  ATTACTCGCGCGCTTCAGCAAAATCCCGCTATTACCGTGGCGAAA 250
   |||||
83  sPheThrSerProValSerGlyLysValValGluIleAsnArgGlyAla 100
   |||||
251  AGCGGTACTTCACTGTCAGTCTGTTGCGGTGAAGCAACGAGGAATC 300
   |||||
100  ysArgValLeuGlnSerValValIleGluValAlaGlyAspAspGlnVal 116
   |||||
301  GAGTTCGAACGCTAGTACCTGAGCGGTGCGCAAAATGAGCAGCGAAA 350
   |||||
117  ThrPheAsnGlyPheGluAlaAsnGlnLeuAlaSerLeuAsnArgAspAl 133
   |||||
351  AGTGGCGCGCAACCTGATTCATCAATCAGCTTATGACCTCGCCTTCGACCC 400
   |||||
133  aIleLysThrGlnLeuValGluSerGlyLeuTrpThrAlaPheArgThrA 150
   |||||
401  GTCCGTTTCAGCAAAATCCCTGCGCTAGATCGCGAGCGTTTCGCATCTTC 450
   |||||
150  rGProPheSerLysValProAlaIleAspSerThrSerGluAlaIlePhe 166
   |||||
451  GTCAATGCGATGGACCAATCCGCTGCTGCGGACCGACCTACGTCATCAT 500
   |||||
167  ValThrAlaMetAspThrAsnProLeuAlaAlaGluProThrValValIle 183
   |||||
501  CAAAGAACCGCGGACGACTTCAACGCGGCTGTTGGTATTGACGCGGC 550
   |||||
183  eAsnGluGlnSerGluAlaPheValAlaGlyLeuAspValLeuSerAla 200
   |||||
551  TGACCGAAGCTAAATCCATGCTGTGAACGACGACGCGGCGGACGTCGC 600
   |||||
200  euThrThrGlyLysValTyValCysLys...LysGlyThrSerLeuPro 215
   |||||
601  TCTGAAATGCTGCCAATATCGAATATTTGGGCGCCGATCC 650
   |||||
216  ArgSerGlnGlnProAsnValGluGluHisValPheAspGlyProHisPr 232
   |||||
651  TGCGCGCTTGAAGGCGCACCATTCATTTCATCGAGCGGCGGCGCA 700
   |||||
232  oAlaGlyLeuAlaGlyThrHisMetHisPheLeuTyProValSerAla 249
   |||||
701  ATAAACCGGTGGACCAATATCAAGACGTGATTCCTATCGGACGT 750
   |||||
249  sPHisValAlaIleTrpSerIleAsnTyrgLnaAspValIleAlaValGlyGln 265
   |||||
751  TTGTTCTGAACGCGCTCTGAATACCGAGCGGTGTTGCTTGGCGGG 800
   |||||
266  LeuPheLeuThrGlyGluLeuTyThrGlnArgValValSerLeuAlaG 282
   |||||
801  CCTGCAAGTCAACAAACCGCGCTCTGCGTACCGTTTGGGTGCGAAGG 850
   |||||

```

```

282  yProValValAsnLysProArgLeuValArgThrValMetGlyAlaSerL 299
851  TGTCTCAACTTACCGCGCGGCAATTTGGTTGACGCGGACAAACCGGTGATT 900
   |||||
299  euGluGlnLeuValAspSerGluIleMetProGlyGluValArgIleLe 315
   |||||
901  TCGGTTTCGGTATTGAACGGTGCAGATTTCGATTCGACAAAGCGCGCATATT 950
   |||||
316  SerGlySerValLeuSerGlyThrLysAlaThrGlyProHisAlaTyLe 332
   |||||
951  GSGAGCTACCAACAATCAGATTTCGTTATTCGAAGAGCGCGCAAG 1000
   |||||
332  uGlyArgTyHisLeuGlnValSerValLeuArgGluGlyArgAspLysG 349
   |||||
1001  AGCTGTTCGGTTCGGTTCGGCGGACGCGGACAAATCTCCATCACGCGC 1050
   |||||
349  luLeuPheGlyTrpAlaMetProGlyLysAsnLysPheSerValThrArg 365
   |||||
1051  ACCACTCTCGGCGCATTTCTCTAAAACAACTCTCAAGTTCACGACAGC 1100
   |||||
366  SerPheLeuGlyHisLeuPheLysGlyGlnValTyAsnMetThrThr 382
   |||||
1101  CGTCAACGCGCGGACGCGCGCATGTTACCGATCGGCGCATTCATGCGCG 1150
   |||||
382  rThrAsnGlySerAspArgSerMetValProIleGlyAsnTyrgLulysV 399
   |||||
1151  TAATCCGCTTGGACATCTCGCTACCTTGTCTTTCGCGCATTTAATTCGTC 1200
   |||||
399  alMetProLeuAspMetGluProThrLeuLeuLeuArgAspLeuCysAla 415
   |||||
1201  GCGCATCCGACAGCGCGCGCGCTTTCGGTTCGTTGGAATTGACGAAAGA 1250
   |||||
416  GlyAspSerAspSerAlaValArgLeuGlyAlaLeuGluLeuAspGluG 432
   |||||
1251  AGACCTCGCTTTCGACGCTCTGCTCCGCGGCAATACGAATACGCGGC 1300
   |||||
432  uAspLeuAlaLeuCysThrPheValCysProGlyLysTyrgLulysGly 449
   |||||
1301  CGCTGTTTCGCGCAAGTCTCGAAACCATTCGAGAGGAAGGC 1341
   |||||
449  lnLeuLeuArgGluCysLeuAspLysIleGluLysGluGly 462

```

seq_name: pir2:H83272

seq_documentation_block:

sodium-translocating NADH dehydrogenase (ubiquinone) (EC 1.6.5.-) nqrl chain PA2999
 C:Species: Pseudomonas aeruginosa
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 16-Feb-2001
 C:Accession: H83272
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
 .; Loty, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
 A:Reference number: AB2950; MUID:20437337
 A:Accession: H83272
 A:Molecule type: DNA
 A:Status: preliminary
 A:Residues: 1-445 <STO>
 A:Cross-references: GB:AE004724; GB:AE004091; NID:g9949083; PIDN:AAG06387.1; GSPDB:GN
 A:Experimental source: strain PA01
 C:Genetics:
 A:Gene: nqrA; PA2999
 C:Superfamily: Vibrio alginolyticus sodium-translocating NADH dehydrogenase (ubiquino
 C:Keywords: oxidoreductase

alignment_scores:

Quality: 1358.00 Length: 447
 Ratio: 3.641 Gaps: 1
 Percent similarity: 83.445 Percent identity: 59.060

alignment_block:

US-09-303-518D-131 x H83272 ..
Align seg 1/1 to: H83272 from: 1 to: 445

282 euAspGluLeuAlaAlaGlyGlnLeuGlnProGlyAsnAsnArgLeuile 298
901 TCCGGTTCGGTATTCACAGCGTCCGATTCACAGCGCGCGCATGATTATTT 950
299 SerGlySerValLeuGlyGlyArgThrAlaHisGlyAlaIyrAlaIyrLe 315
951 GGCAGCTTACCAATCAGATTTCGGTATTCGAAAGAGCGCGCAAG 1000
315 uGlyArgTyrHisLeuGlnLeuSerCysLeuLysGluGlyAspGlnArg 332
1001 AGCTGTCGGTGGGTTCGGCGCGCGGACAAATCTCAAGTTTCAGCAGC 1050
332 LuPheLeuHisTyrLeuArgAlaGlyValGluLysHisSerLeuLeuAsn 348
1051 ACCATCTCGGCCATTTCCTTAAACAACTCTTCAAGTTTCAGCAGC 1100
349 ValPheValSerArgLeuLeuGlyGlyLysArgPheAlaPheThrSe 365
1101 CGTCAACGGCGCGCGCGCATGTACCGATCGGCACGCTTATGAGCGG 1150
365 rThrAsnGlySerProArgAlaMetValProValGlyAsnTyrGluAlaV 382
1151 TAATGCGGTTCGACATCCTGCTACCTGCTTTCGCGCATTAATCGTC 1200
382 aLMetProLeuAspIleLeuProThrGlnLeuLeuArgTyrLeuileVal 398
1201 GCGGATACCGACGCGCGCGCTTGGTGTCTTGGAAATTTGGACGAAGA 1250
399 GlyAspThrGluMetAlaGlnLysLeuGlyAlaLeuGluLeuAspGluG 415
1251 AGACCTCGCTTTGTGAGCTTCGTCGCCGGGCAATACGAATACGCGCC 1300
415 uAspLeuAlaLeuCysSerTyrValCysAlaGlyLysTyrGluTyrGlyP 432
1301 CGCTGTTCGCAAGCTGCTGGAACCATTTGAGAAGGAAGGC 1341
432 rotLeuArgAspAsnLeuAlaArgIleGluGlnGly 445

seq_name: pir2:C81751

seq_documentation_block:
probable sodium-translocating NADH dehydrogenase (ubiquinone) (EC 1.6.5.-) alpha chain
C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 16-Feb-2001
C:Accession: C81751
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39
A:Reference number: AB1500; MUID:20150255
A:Accession: C81751
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-465 <TET>
A:Cross-references: GB:AE002269; GB:AE002160; NID:g7190041; PIDN:AAF38896.1; PID:g719
A:Experimental source: strain Nigg (MoPn)
C:Genetics:
A:Gene: TC0002
C:Superfamily: Vibrio alginolyticus sodium-translocating NADH dehydrogenase (ubiquino
C:Keywords: oxidoreductase

alignment_scores:
Quality: 453.00 Length: 464
Ratio: 1.520 Gaps: 14
Percent Similarity: 64.224 Percent Identity: 28.448

alignment_block:
US-09-303-518D-131 x C81751 ..
Align seg 1/1 to: C81751 from: 1 to: 465

1 ATGATTAAATCAAAAAAGGTCTAATCTGCCATCGCGCGCGACACCGGA 50
1 MetIleLysIleuArgGlyLeuAspLeuProLysSerGlyAlaProG 17
51 GCAAGTCATTTATGACGGCGCGCGCATACCGAAAGTCGCGTGTGGCG 100
17 uGlnArgIleGluAlaAlaArgProValArgSerValAlaLeuileGlyP 34
101 AGAATATCTCGCATCGCGCGCGCGCTCGATCAAAATCAAGGAAGTCAAGCC 150
34 heAspTyrHisGlyMetLysProThrMetAlaValGlnValGlyAspArg 50
151 GTCAAAAAGGCCAAGTCGTGTTTGAAGACAAAAAGATCCGGCGTAGT 200
51 ValLysLeuGlyGlnValLeuPheThrAspLysLysAsnProSerValSe 67
201 ATTACTGCGCGCGCTTCAGGCAAAATCGCGCTATTACCGTGGCGAAA 250
67 rTyrThrAlaProGlyAlaGlyValValSerAlaIleHisArgGlyGluL 84
251 AGCGGTACTTACGTACGTGCTGCGTTCGCTGGAAGGCAACGACGAAATC 300
84 ysArgValLeuGlnSerValIleAspLeuAspGlyAspGluGlnLeu 100
301 GAGTTCGAACGCTTACCTGACGCGCTGCGCAAAATTTGAGCAGCGAAA 350
101 GluPheAlaArgTyrProAlaAspLysLeuAlaThrLeuSerAlaGluG 117
351 AGTGGCGCGCAACTGATTCATCAGCTTATGAGCTGCGCTTCGACCC 400
117 nValArgAspAsnLeuGlnSerGlyLeuThrAlaLeuArgThrA 134
401 GTCGTTCACAAATCCCTGCGGTAGATGCGCGAGCGCTTCGCCATCTTC 450
134 rGProPheSerLysValProAspProGluSerSerProSerIlePhe 150
451 GTCATTCGATGACACCAATCGCTGCGTTCGCGACCTACGCTCATCAT 500
151 ValThrAlaIleAspThrGlnProLeuAlaAlaAspProGlnValIle 167
501 CAAGAAGCCCGCGAAGCTTCAACGCGCGCTGTTGTTGATGAGCGCG 550
167 eAlaGluGlnGlyGluAlaPheGlnAlaGlyLeuThrValLeuGlyArgL 184
551 TGACCGAAGCTTAAATCCATGTGTGTAAAGCAGCAGCGCGCAGCGCGG 600
184 euAla.....ArgValPheLeuCysLysAlaGluGlyValSerLeuPro 198
601 TCTGAAATGCTGCCATATCGAAACACATGAATTTGGCGCGCGCATCC 650
199 GlyGluAlaLeuSerGlyValThrAlaGlnAlaPheSerGlyProHisr 215
651 TGCGCGTTGAGTCGCGCGCATTCATTTTCATCGCCAGCTCGCGCGGA 700
215 oAlaGlyLeuProGlyThrHisIleHisPheLeuAspProValGlyAlaG 232
701 ATAAACCGGTGGACCATCAATATCAACAGCTGATGCTATCGACGCT 750
232 lLysSerValTrpAsnLeuAsnTyrGlnAspValIleAlaIleGlyLys 248
751 TTGTTCGTAACAGCGCGTCTCAATACGAGCGGTGCTGCTTGGCGCGG 800
249 LeuPheThrThrGlyGlnLeuThrPheGluArgValIleAlaLeuAlaG 265
801 CCTGCAAGTCAACAAACCGCGCTCTTCGCTACCGTTCGTTGGTGGCGAAG 850
265 yProValValGluLysProArgLeuValArgThrArgLeuGlyAlaAsnL 282
851 TGCTCTCACTTACCGCGCGCGAATTTGTTGACGCGGACACCGCGTGTAT 900

— — — — —

3 IleThrValAsnArgGlyLeuLeuSerLeuLeuGlnGlySerProLysGlu 19
51 GCAAGTCATTATGACGGCCGGCCCATACCGAAGTCGGTTCGTTGGCG 100
19 uSerGlyPheTyrAsn.....LysIleAsp 28
101 AGAATATGTCGC.....ATCGGCCCC.....TCG 126
28 roGluPheValSerIleAspLeuArgProPheGlnProLeuSerLeuLys 44
127 ATGAAATCAAGGAGGTGAAGCGTCAAAAAGGCGCAAGTGTGTTGA 176
45 LeuLysValGlnGlnGlyAspAlaValCysSerGlyAlaProIleAlaGlu 61
177 AGCAAAAAGATCCGGCGCTAGTATTACTGCGCGCGCTTACGGCAAAA 226
61 uTyrLysHisPheProAsnThrTyrIleThrSerHisValSerGlyVal 78
227 TCGCGCTATTACCTGCGGAAAGCGGTACTTCCAGTCAGTCGTGATT 276
78 alThrAlaIleArgGlyAsnLysArgSerLeuLeuAspValIleIle 94
277 ...CGCGTTGAAGCAACGACGAAATCGAGTTCGAACGCTACCTGA 323
95 LysLysThrProGlyProThrSerThrGluTyr...ThrTyrAspLeuGlu 110
324 AGCGTGGCAAAATTCAGCAGCAAAAAGTGCGCCCAACCTGATTCAAT 373
110 nThrLeuSerArgSerAspLeuSerGluIlePheLys.....GluA 124
374 CAGCGTTATGAGTTCGCTTCGCAACCGCTCGCTTCAGCAAAAATCCCTGCC 423
124 snGlyLeuPheAlaLeuIleLysGlnArgProPheAsp...IleProAla 139
424 GTAGATGCCGAG...CGTTTCGCATCTTCGTCAATGCGATGACACCAA 470
140 IleProThrGlnThrProArgAspValPheIleAsnLeuAlaAspAsnAr 156
471 TCGCGTGGCTGCGACCTACGCTCATCATC.....A 502
156 gProPheThrProSerProGluLysHisLeuAlaLeuPheSerSerArg 173
503 AAGAACCGCGCAGAGACTTCAACCGCGCTGTGTGTTATTGACCGCGCTG 552
173 luGluGlyPheTyrValPheValValGlyValArgAlaIleAlaLysLeu 189
553 ACCGAAGTAAATCCATGTGTGTAAAGCAGCAGCGCCACAGCTGCGCTC 602
190 PheGlyLeuArgProHisIleValPheArgAspArgLeuThrLeuProTh 206
603 TGAATATGCTGCCAATATC...GAAACACATGAATTTGCGCGCGCGCATC 649
206 rGlnGluLeuLysThrIleAlaHisLeuHisThrValSerGlyProPheP 223
650 CTGCGCGCTGAGTGGCAGCAGCATTCATTCATCGAGCCAGCTGCGCGCG 699
223 roSerGlySerProSerIleHisIleHisSerValAlaProIleThrAsn 239
700 AATAAA...ACCGTGTGACCATCAATATCAAGACGTGATTGCTATCGG 746
240 GluLysGluValValPheThrLeuSerPheGlnAspValLeuThrIleGlu 256
747 ACGTTTGTGTGAACAGCGCTCTGAATACCGCGCGCTGTGTCCTGG 796
256 yHisLeuPheLeuLysGlyArgIleLeuHisGluGlnValThrAlaLeuA 273
797 CGCGC.....CTGCAAGTCAACAAACCGCGCTCTTTCGCTACCGTTTG 840
273 laGlyThrAlaLeuLysSerSerLeuArgArgTyrValIleThrLys 289
841 GGTGCGAGGTGCTCAACTTACCGCGCGCAATTTGTTGACGGGACAA 890
290 GlyAlaSerPheSerSerLeuAsnLeuAsnAspIleSerAspAsnAs 306

891 CCGCGTGATTTCCGGTTCGGTATTGAACGTTGCGATTGACACAGGCGCG 940
306 pThrLeuSerGlyAspProLeuThrGlyArgLeuGlyCysLysLysGluG 323
941 ATGAT...TATTGGGAGCGTTACACCAATCAGATTTCCGTTATCGAAGA 987
323 luGluProPheLeuGlyPheArgAspHisSerIleSerValLeuHisAsn 339
988 GCGCGCAGCAAGAGCTGTTCCGCTGGTTCGCGCGCGCAGCGGACAAAT 1037
340 ProThrLysArgGluLeuPheSerPheLeuArgIleGlyPheAsnLysPr 356
1038 CTCCATCAGCGCACCTCTCGGCCATTTCTTAAACAACTCTTCA 1087
356 oThrPheThrLysThrThrLeuSerGlyPhePheLysLysArg...T 372
1088 AGTTTCACG.....ACAGCGTCAACGCGCGCGCGCGCATCGTA 1128
372 hrTyrThrAsnProAspThrAsnLeuHisGlyGluThrArgProIleIle 388
1129 CCGATCGCACTTATGACGCGTAATCGCGTGGACATCCTCGCTACCTT 1178
389 AspThrAspIleTyrAspLysValMetProMetArgIleProValValPr 405
1179 GCTTTTGGCGGATTTAATCGTCGCGATACCGCAGCGCGCGCTTGG 1228
405 OleuIleLysAlaValIleThrLysAsnPheAspLeuAlaAsnGluLeuG 422
1229 GTTGCTTGAATTTGACGAAGAACCTCGCTTTGTCGACGCTTCGCTGCG 1278
422 lPheLeuGluValCysGlyGluAspPheAlaLeuProThrLeuIleAsp 438
1279 CGGGAATAATACGATACGCGCTGTGCGCAAGTGTGGAACCAT 1328
439 ProSerLysThrGluMetLeuThrIleValLysGluSerLeuIleGluTy 455
1329 TGAGAAGGAA 1338
455 rAlaLysGlu 458
seq_name: pir2:E72040

seq_documentation_block:
probable sodium-translocating NADH dehydrogenase (ubiquinone) (EC 1.6.5.-) alpha chain
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 16-Feb-2001
C:Accession: E72040; G81623
R:Kelman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, R.; Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; MUID:99206606
A:Accession: E72040
A:Molecule type: DNA
A:Residues: 1-467 <ARN>
A:Cross-references: GB:AE001656; GB:AE001363; NID:g4377047; PIDN:AAD18882.1; PID:g437
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39
A:Reference number: A81500; MUID:20150255
A:Accession: G81623
A:Molecule type: DNA
A:Residues: 1-467 <REA>
A:Cross-references: GB:AE002164; GB:AE002161; NID:g7188939; PIDN:AAF37899.1; PID:g718
A:Experimental source: strain AR39, HL cells
C:Genetics:
A:Gene: nqrA; CP0002
C:Superfamily: Vibrio alginolyticus sodium-translocating NADH dehydrogenase (ubiquinol)
C:Keywords: oxidoreductase

A;Gene: VC1015

A:Gene: VC1015

A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-673 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD01910.1; PID:g16502752; GSPDB:GN00176
C:Genetics:
C:Gene: STV1665
C:Superfamily: unassigned ferredoxin 2[4Fe-4S]-related proteins; ferredoxin 2[4Fe-4S] h
alignment_scores:
Quality: 151.50 Length: 449
Ratio: 0.728 Gaps: 17
Percent Similarity: 46.325 Percent Identity: 22.049
alignment_block:
US-09-303-518D-131 x AE0692 ..
Align seg 1/1 to: AE0692 from: 1 to: 673
100 GAAGAAATATGTCGGCATCGCCCTCGATGAAATCAAGGAAGGTGAAGC 149
.....
48 LysGlnHisIleGlyAlaGluGlyGluLeuCysValSerValGlyAspArg 64
.....
150 CGTCAAAAGGCAAGTCTGTTGAAGACAAAGAAAGATCCGGCGTAG 199
|||
64 gValLeuArgGlyGlnAlaLeuThrArgGlyArgGlyArgMetLeuProV 81
.....
200 TATTACTGCGCGGCTTCAGGCAAAATCGCCGCTATT 237
|||
81 alHis...AlaProThrSerGlyThrValIleAlaIleAlaProHisSer 96
.....
238CACGTCGGCGAAAGCGCTACTTCAGTCAGTCGTCGATTCGCGT 281
|||
97 ThrAlaHisProSerAlaLeuAlaGluLeu...SerValIleIleAspAl 112
.....
282 TGAAGGCAACGACGAAATCGAGTTTCGAAC 310
.....
112 aAspGlyGluAspArgTrpIleGluArgGluGlyTrpSerAspTyrArgA 129
.....
311 GCTACGTACCTGAAGCGCTGCAAAATGAGCAGCGAAAGATGCGCCGC 360
|||
129 laHisSerArgGluAlaLeuIleGluArgIleHisGlnTyrGlyValAla 145
.....
361 AACCTGATTCAATCAGGCTTATGACTGCGCTTCGACCGCTCCGTCAG 410
|||
146 GlyLeuGlyGlyAlaGlyPheProThrGlyVal 156
.....
411 CAAATCCCTGCGCTAGATCCGAGCGGTTCGCCATCTTCGTCATCGCA 460
|||
157 .LysLeuGlnGlyGlyAspAspLysIleThrThrLeuIleIleAsnAla 173
.....
461 TGGACACCAATCGCTGCTGCGGACCCCTACGCTCATCATCAAGGAAGCC 510
|||
173 laGluCysGluProTyrIleThrAlaAspArgLeuMetGlnAspCys 189
.....
511 GCCGAGAGACTCAACAGCGGCTGTTGGTATTGAGCGCGCTGACCGAA.. 558
|||
190 AlaAlaGlnIleValGluGlyIleArgIleLeuAlaHisIleLeuGlnPr 206
.....
559CGTAAATCCATG 571
|||
206 oArgGluValLeuIleGlyIleGluAspAsnLysProGlnAlaIleSerM 223
.....
572 TGTCTAAGCAGCAGCGCGCTGCGCTGTAATAATGCTGCAATATC 621
|||
223 etLeuArgAlaValLeuAlaAspAlaHisAspIle 234
.....
622 GAAACACATGAATTTGGCGCGCCGCTCCGCGGC 657
|||
235 SerLeuArgValIleProThrLysTyrProSerGlyGlyAlaLysGlnLe 251
.....
658TTGAGTCGACGCGACATTCATTTCATCGACGCGCGCG 697
|||
251 uThrGlnIleLeuThrGlyLysGlnValProHisGly 264
.....
698 CGAATAAAGCGGTGTGACCATCAATTTATCAAGACGTGATT 738
|||
264 lYArgSerSerAspIleGlyValLeuMetGlnAsnValGlyThrAlaTyr 280
.....
739 GCTATCGAGCGTTCGTTGTAACAGCGCGCTGAATACCGAGCGGTGT 788
|||
281 AlaValLysArgAlaValValAspGlyGluProIleThrGluArgValVa 297
.....
789 TGCCTTGGCGGCGCTCAAGTCAACAAACCGCGCTCTTCGTCACCGTT 838
|||
297 lThrLeuThrGlyGluAlaIleSerArgProGlyAsnValTrpAlaArgL 314
.....
839 TGGGTGCGAAGGTGCTCAACTTACCGCGCGCGAATTGGTTGAC 882
|||
314 euGlyThrProValArgHisLeuLeuAsnAspAlaGlyPheCysProser 330
.....
883 CGGACAAACCGCGGTGATTTCGGTTCGGTATTGAACGTCGATTGCAC 932
|||
331 AlaAspGlnMetValIleMetGlyGlyProLeuMetGlyPheThrLeuPr 347
.....
933 AGGCGCGCATGATTATTGGGACGCTACCAATCAGATTT 973
|||
347 oTrpLeuAspValProValValLysIleThrAsnCysLeuLeuAlaProS 364
.....
974 CCGTTATCGAAAGAGCGCGCAAGAGAGCTGTTCGGCTGGGTGGCGCG 1023
|||
364 erValThrGluMetGlyAlaPro 371
.....
1024 CAGCGGACAAATACTCCATCAGCGGACCACTCTCGGCCATTTCTTAA 1073
|||
372 GlnGluGluLysSerCysIleArgCysSerAlaCysAlaAsp 385
.....
1074 AAACAAACTCTTCAAGTTACAGCAGCGCTCAAGCGGCGCGCGCGCCA 1123
.....
385
1124 TGGTACCGATCGGCATTTATGAGCGGTAAATGCGGTGGACATCTGCCT 1173
|||
386AlaCysProAlaAspLeuPro 393
.....
1174 ACCTGCTTTCCGCGATTAACTGTCGGCATACCGACAGCGCGCAGGC 1223
|||
394 GlnGlnLeuTyrTrpPheSerLysGlyGlnGlnHisAspLysAlaThrAl 410
.....
1224 TTTGGTTCGTTGGAATTGGACGAAGAGACCTCGCTTGT 1264
|||
410 ahHisIleAlaAspCysIleGluCysGlyAlaC 422
.....
1265 GCAGCTTCGTCGCGCGCAATACGAATACGCGCGCTGTCGCGC 1311
|||
422 ysAlaTrpValCysProSerAsnIleProLeuValGln 434
seq_name: pir2:S59310
seq_documentation_block:
probable membrane protein YMR317w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein YMR924.09
C:Species: Saccharomyces cerevisiae
C:Date: 29-Nov-1995 #sequence_revision 23-Feb-1996 #text_change 05-Nov-1999
C:Accession: S59310
R:Churcher, C.M.
submitted to the EMBL Data Library, September 1995
A:Reference number: S59302
A:Accession: S59310
A:Molecule type: DNA
A:Residues: 1-1104 <CHU>
A:Cross-references: EMBL:Z54141; GSPDB:GN00013; MIPS:YMR317w
A:Experimental source: strain AB972
C:Genetics:
A:Gene: MIPS:YMR317w
A:Map position: 13R

alignment_scores:

Quality: 143.00 Length: 480
Ratio: 0.633 Gaps: 17
Percent Similarity: 47.083 Percent Identity: 22.292

alignment_block:

US-09-303-518D-131/rev x S59310

Align seg 1/1 to: S59310 from: 1 to: 1104

```
1331 TCATGCTTCCAGCATTTGGCCAAACAGCGCGCGTATTCGTATTCGCC 1282
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
350 SerSerThrSerSerValSerSerGluAlaProSerSer..... 363
1281 CGGGCAGACGAAGCTGCACAAAGCGAGGTCTTCTTCGTCCTCAAGC 1232
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
364 .....ThrSerSerValSerSerGluA 372
1231 AACCCAAAGCCTGCGCGTGTGCGGTATGCGCGACGATTAATCGCGCAA 1182
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
372 laProSerSerThrSerSerValSerSerGluAlaSerSerThrLys 388
1181 AGCAAGGTAGGCAGATGTCCAAACGCGATTACGCGCTCATAGTCCGAT 1132
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
389 SerSerVal.....MetSerSerGluValSerSerAla.....Th 400
1131 CGGTACCATGGCGCGTGTGCGCGCGTTCGACGGCTGCTGCACTGAAGA 1082
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
400 rSerSerLeuValSerSerGluAlaProSerAlaIleSerSerLeuAlaS 417
1081 GT.....TTGTTTTAGGAATGGCGAGAGTGGTGGCGGTGATG... 1041
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
417 erSerArgLeuPheSerSerLysAnThrSerValThrSerThrLeuVal 433
1040 .....GAGTATTTTCGCGTGGCGCGCAACCCAGCGACAGCTCTTT 997
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
434 AlaThrGluAlaSerSerValThrSerSerLeuArgProSerSerGluTh 450
996 GCTGGCGCTTCTTGATPACGGAATCTGATTTGTGTAGCGTCCCAAT 947
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
450 rLeuAlaSerAsnSerIleGlu..... 458
946 AATCATGCGCCTGTGTCATCGCACCGCTCAATACGGAACCGGAATC 897
458 ..... 458
896 ACGCGTGTCCGGTCAACCAATTGCGCGCGGTAAAGTTGAGACACCTT 847
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
459 SerSerLeuSerThrGlyTyAsnSerThrValSerThrThrThrSerAl 475
846 CGCACCAAAACGTAGCAAGAGCGCGGTGTTGTTGACTTGCAGCGCGC 797
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
475 aAlaSerSerThrLeuGlySerLysValSerSerAsnSerArgMeta 492
796 CCAAGGCAACACCGCGTCTGCTATTCAGACGGCTGTACCAACAACT 747
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
492 laThrSerLysThrSerSer..... 498
746 CCGATAGCATTCAGCTTGTGATAATTGATGTCACACGGTATTATTCGC 697
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
499 .....ThrSerSerAspLeuSerLysSerSerValIlePheGl 511
696 GCGGACTGGCTCGATGAATGAATGTGCGTCCCACTCAACCGCGCAGGAT 647
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
511 yAsnSerSerThrVal.....T 517
646 CGCGGCGCGCAAAATTCATGTGTTTCATATTGGCAGCA..... 609
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
517 hrThrSerProSerAlaSerIleSerLeuThrAlaSerProLeuProSer 533
608 ...TTTTCAGACGCGCAGCTGCGCGCTGCTGCTTTACACACATGGATTTT 562
```

```
534 ValTrpSerAspIleThrSerSerGluAlaSerSerIleSerSerAsnLe 550
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
561 ACGTTGCGTCACGGCGCTCAATACCAACAGCGCGCGTTTGAAGTCTTCGG 512
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
550 uAlaSerSerSerAlaProSerAspAsnAsnSerThrIleAlaSer.... 565
511 CGGCTTCTTTGATGATGACC.....GTAGGTCGCGCA 480
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
566 ..AlaSerLeuIleValThrLysThrLysAsnSerValValSerIle 581
479 GCCACGCGATTGCTTCATCGCATTCACGAAGATGGCGAACGCTTCGGC 430
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
582 ValSerIleThrSerSerSerGluThrThrAsnGluSerAsnLeuAlaTh 598
429 ATCTACGCGAGGATTTTGTCTGAAC..... 405
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
598 rSerSerThrSerLeuLeuSerAsnLysAlaThrAlaArgSerLeuSerT 615
404 ..GACGCGTCCGAAGCGCATTCATAGCCTGATTGAATCAGGTTCGCG 357
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
615 hrSerAsnAlaThrSerAlaSerAsnValPro.....Thr 626
356 CGCACTTTTTCGCTGCTCAAT.....TT 334
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
627 GlyThrPheSerSerMetSerSerHisThrSerValIleThrProGlyPh 643
333 TGCAGCGCTTCAGGTACGTAGCTTTCGAACTCGATT.....TCGT 293
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
643 eSerThrSerSerAlaSerLeuAlaIleAsnSerThrValValSerSerS 660
292 CGTTGCTTCAACGGCAATCACGACTCAGTCAAGTACGCGCTTTTCGCCA 243
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
560 erLeuAlaGlyTySerPheSerThrProGluSerSerProThrThrSer 676
242 CGGTGAATACGCGGATTTTGCCTGAAGCGCGCAGTAAATACACGCC 193
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
677 ThrLeuValThrSerGluAlaProSerThrValSerSerMetThrThrSe 693
192 CGGATTTCTTTTGTCT.....TCAACACGACACTT 164
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
693 rAlaProPheIleAsnAsnSerThrSerAlaArgProSerProSerThra 710
163 GGCGTTTTTTCAGCGCTTCACCTTCTCTGATT..... 132
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
710 laSerPheIleThrGluSerThrSerIleSerSerValProLeuAla 726
131 .....TTCATCGAGGGCGCATGCCGACATATTC 103
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
727 SerGlyAspValThrSerSerLeuAlaAlaHisAsnLeuThrThrPheSe 743
102 TTCGCCAACGACGCGACTTCGGTATATGCGCGCGCGTCA 63
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
743 rAlaProSerThrSerSerAlaGlnLeuValSerLysSer 756
```

seq_name: p1r2.T02345

seq_documentation_block:

hypothetical protein KIAA0324 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 05-Nov-1999
C:Accession: T02345
R:Ricke, D.O.; Bruce, D.; Mundt, M.; Doggett, N.; Munk, C.; Saunders, E.; Robinson, D.
re, J.; White, S.; Ueng, S.; Tatum, O.; Campbell, C.; Fawcett, J.; Deaven, L.
submitted to the EMBL Data Library, March 1998
A:Description: Sequencing of human chromosome 16p13.3.
A:Reference number: Z14664
A:Accession: T02345
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1791 <RIC>
A:Cross-references: EMBL:AC004493; NID:g2996648; PIDN:AAC08453.1; PID:g2996650
C:Genetics:

CGTGGACCACTTATCAAGACGTCGTTCTATCGGACGCTTCTCTG
|||||:|:|

```
seq_documentation_block:
rnc protein homolog - Haemophilus influenzae (strain Rd KW20)
C:Species: Haemophilus influenzae
C>Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change
CiAccession: E64136
R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirk,
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.;
D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.
A:title: Whole-genome random sequencing and assembly of Haemophilus
A:Reference number: A64000; MUID:95350630
A:Accession: E64136
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-819 <TIGR>
A:Cross-references: GB:U32841; GB:I42023; NID:gl574529; PIDN:AAQ
C:Superfamily: unassigned ferredoxin [24Fe-4S]-related proteins.
E:371-435/domain: ferredoxin [24Fe-4S] homology <PDB>
```


alignment_scores:

```

alignment_block:
US-09-303-518D-131 x E64136 ..
Align seg 1/1 to: E64136 from: 1 to: 819

Quality: 135.50 Length: 465
Ratio: 0.622 Gaps: 15
Percent Similarity: 46.882 Percent Identity: 19.570

```

70 CCGGCCATTACCGAAGTCGCGTTTGGCTGTGGCGAAGAATATGTCCGCATCGC
||| |||||::: ||| ::::|||||
38 ProLeuGlyThrAspPheTyrlleProLeuLysGlnHisLeuGlyThrH 54

120 CCCCTCATTAANCAACGAAGGTGAACCGGTCAAAAAGCCCAAGTC 169
: ::::: |||||||||::: ||| :|||:
54 rGlyAsnLeuIleLysGluGlyAspTyrrValLeuLysGlyGlnAla 71

170 TGTTTTGACACAAAAAGATCCGGCGTAGTTATTACTCGCGCGCTTCA 219
|| ::::: ||| :|||:
71 euThrLysGlyAspGlyLeuArgMetLeuProValHisAlaProThrSer 87

220 GCGAAATGCGCGCTATTACC CGTGGCGAAAGCGCGTACTTTCAGTCAGT 269
||| ||| ||| ||| :|||:
88 GlyThrIle.....LysSerIleLysProTyrrVa 97

270 CTGTATTCGGTTGAAGCACGACGAATCCAG..... 303
||::: ||||| |||||
97 IalaThrHisProSerGlyLeuAspGluProThrlleHisLeuGlnAlaa 114

304TCCAACCTACGTACCTGAAGCGCTGCCAAAA 336
:|||||
114 spGlyLeuAspGlnTrpIleGluArgsnProIleAspSpheSerThr 130

337 TTGACGACGGAAGTCGCGGCAACCTGATTCAATCAGGCTTATGAC 386
|||||||::: ||| :|||:
131 LeuSerSerGluGlnLeuIleHisLysIleTyrrGlnAlaGlyIle..Al 146

387 TGCGCTTCGCAACCGTCGCTC.....AGCAAAATCCCTCGCGTAG 427
: ||||| ||| :|||:
146 aGlyLeuGlyGlyAlaValPheProThrAlaAlaLysIleGlnSerAlag 163

428 ATGCGAGCGCTTCGCCATCTTCGTCATCGGATGGACACCAATCCGCTG 477
::: |||||::: ||| :|||:
163 luGlnLysValLysLeuLeuIleIleasnGlyAlaGluCysGluProTyrr 179

478 GCTCCGACCTCAGTGCATCATCAAGAAGCCGCGGAAGACTTCAAACG 527
::: |||||::: ||| :|||:
180 IleThrCysAspAspArgLeuMetArgGluArgAlaAspGluIleIleLy 196

528 CGGCGCTGTGTTAGCCGCTTCAGCGACGAGTAAATCCATCTGTGTA 577
|||::: |||||::: ||| :|||:
196 sGlyleArgileLeuArgTyrrIleLeuHisProGlu...LysValValI 212

578 AAGCAGGCGCGACAGCTGCCGTCT..... 603
||| ::: |||||::: ||| :|||:
212 leAlalleGluAspAsnLysProgluAlalleSerAlalleArgAsnAla 228

604 ...GAAATCGCTCCATATCGAACACATGAATTTGGCGGCGCGCATPC 650
||| ||| :|||::: ||| :|||:
229 LeuGlnGlyAlaAsnAspIleSerIleArgValIleProThrLysTyrrPr 245

651 TGCGCGCTCAGTGGCAGCACATTCATTCATCGAG.....C 688
|||::: |||||::: ||| :|||:
245 oSerGlyAlaThrLysglnLeuIleTyrrLeuLeuThrGlyllecluValP 262

689 CAGTCGCGCGGATAAAACCGTGTGGACCATCAATATCAAGACGTG... 735
|| ||| :|||::: ||| :|||:
262 roSerGlyGluArgSerSerSerIleGlyValLeuMetGlnAsnValcly 278

736ATTCTATCGGACGTTTTTCGTAAACAGGCGCTGAATACCGA 779
||| ||||| ||| :|||:
279 ThrMetPheAlalleLysArgAlalleAsnAspGluProLeulleGI 295

780 GCGCGTGTTCCTTGCCGCGCCTGCAGTCAACAACACGGCGCCTCTTGC 829
||||| ||||| ||||| |||||
295 uArgValThrLeuThrGlyAsnLysIleAlaGluLysGlyAsnTrpT 312
:::
830 GTACCGTTTTGGGTGCGAAGGTGTCtCAATT.....ACGCCGGCGAA 873
||||| ||||| :::
312 rpValArgLeuGlyThrProIleSerGlnIleLeuSerAspAlaGlyTr 328
|||
874 TTGGTTGACGCGGACAAACC CGTgATTTCGGGTTCGGTAATTAACCGTGC 923
||| ||||| ||||| :::
329 GlnPheAspLysHisPheProIlePheAlaGlyGlyPrometMetGlyLe 345
924 GATTGCACAGGCCGCGCATGATTATTGGGACGCTACCACATCAGATT 973
345 u.....:::
GlutLeuP 348
974 CCGTATTCSGAAGAGCGCGAGCAAGAAGAGTGTTCGGCTGGGTTCGCCG 1023
::::: ||||| :::
348 roAsnLeuAsnAlaProValThrLysLeuValAsnCysLeuLeuAlaPro 364
1024 CAGCGGACAAATACTCCATCACGCGCACCACTCTCGGCcATTTCCTAAA 1073
::: |||||
365 AspTyrLeuGluTrp.....369
1074 AAACAACTCTCAAGTTCAGCAGACGCGCTCAACGGCGCGGACGCGCCCA 1123
||| ::::: |||||
370AlaGluProGluAlaGluGlnAlaC 378
1124 TGCTACCGATCGGCACATTATGACGGCGTATGCGTTGGACATCTCGCCT 1173
::: ::::: |||||
378 ystIleArgCysSerCysSerAspAlaCysProValAsnLeuMetPro 394
1174 ACCTTGCTTTGGCGGATTTAATCTGCGGGCATACGCACAGCGCGCAGGC 1223
||| ||| ::::: |||||
395 GlnGluLeuTyrTrpPheAlaArgSerGluAspHisLysLysSerGluGI 411
1224 TTTGGGT.....TGCTTGAATTGGCAGGAAGACCTCGCTTTGT 1264
::: ||||| |||||
411 uTyrAlaLeuLysAspCysileGlu.....CysGlyIleC 423
1265 GCAGCTTCGCTGCCCGCAAATACGAATACGCCCGCTGTTGGCG... 1311
||||| ||||| |||||
423 ysAlatyrValCysPro.....SerHisIleProLeuIleGlnTrp 436
1312AAAGTCGTGGAAACCATTGAGAAGAA 1338
||||| :::
437 PheArgGlnGluLysAlaylsIleTrpGlnIleLysGluLysGln 451

seq_name: pir2:E72398

seq_documentation block:

seq_documentation_block:
hypothetical protein TM0244 - Thermotoga maritima /strain msp09

C;Species: Thermotoga maritima

```
C; Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
```

C;Accession: E72398

R; Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey, J.R.; Eide, M.J.; Arnold, J.F.; Peterson, D.J.; White, T.L.; Garret, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, C.M.

C.M.
Nature 399, 323-329, 1999

A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genomes

A;Reference number: A72200; MUID:99287316

A;Accession: E72398

A;Status: preliminary

A;Molecule type: DNA
A:Residues: 1-151

A;Residues: 1-451 <ARN>
A:Cross-references: CB:

A;CROSS=References: GB:AE001708; GB:AE000512; NID:g4980740; PIDN:AAD35335.1; PID:g4980740
A;Experimental source: strain MSR8

C;Genetics:

A;Gene: TM0244

[illegible]

Quality: 135.00 Length: 444
Ratio: 0.603 Gaps: 17
Percent Similarity: 50.450 Percent Identity: 20.270

alignment_block:

US-09-303-518D-131 x E72398 ..

Align seg 1/1 to: E72398 from: 1 to: 451

```

49 GAGCAAGTCATTATGACGGCGCGCCATTACCGAAGTCGGTTCGTTGG 98
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
35 AspLysProIleGluArgAlaProLeuProGlnLysValPheValPheLe 51
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
99 CGAAGATATGTCGCATCGCGCCCTCGATCAATCAAGGAAGTGAAG 148
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
51 uSerAsnHisAlaGlyAsnProAlaLysProValValSerProGlyAsp 68
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
149 CCGTCAAAAAGGCCAAGTGGTGTGAAGACAAAAGAACCCGGCGCTA 198
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
68 luValLysThrGlyGlnValIleGlyGluProGluGlyPheIleSerAla 84
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
199 GTATTACTGCGCGCTTCAGCAAAATCGCGCTATTACCGTGGCGGA 248
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
85 TyrLeuHisSerProValThrGlyArgValLeuGluIle..... 97
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
249 AAGCGCGTACTTCAGTCAGTCGTCG.....ATTGCGG 280
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
98 LysLysIleLeuHisProIleLeuGlyLysProIleGluAlaIleValI 114
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
281 TTGAAGGCAACGACGAAATCGAGTTCGACGCTAGCTCAAGCGCTG 330
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
114 leGluArgThrSerAspAspGluTrpValHisIleGluThrGlyAspPhe 130
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
331 GCAAATTCGAGCGCAAAATCGCGCCACCTGATTCATCA..... 375
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
131 GluArgMetSerLysGluIleLeuGluIleLysLysAlaGlyIle 147
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
376 ....GCTTATGACGTCGCTTCGACCGCTCGCTTCAGCAAAATCCCTG 421
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
147 eValGlyLeuGlyGlyAlaMetPheProThrHisValLysLeuSerProp 164
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
422 CGTAGATCGGAGCGGTCGCGCATTCGTCGAATCGGATGACACCAAT 471
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
164 roProGluLysLysValAspThrLeuIleValAsnGlyAlaGluCysGlu 180
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
472 CCCTGCTGCGGACCTACGGTCATCATCAAGAACGCGCGAAGACTT 521
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
181 ProValLeuThrIleAspHisArgLeuMetLeuGluArgAlaGluAspI 197
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
522 CAACGCGGCTGTTGGTATTGAGCGCGCTGACCGCAACGTAATCCATG 571
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
197 eLeuGlnGlyIleLeuIleMetMetLysVal.....LeuGlyV 210
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
572 TGTGTAAGCA...GCAGGCGCAGAGTCGCGCTCTGAAATGCTGCCAAT 618
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
210 alGlnLysAlaValValGlyValGluSerAsnLysMetAspAlaTyrHis 226
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
619 ATCGAAACACATGAATTTGCGGCGCGCATTCCTGCGGCTTGAGTGGCAC 668
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
227 AsnLeuLysLysValPheLysGly...TyrProValAspValAlaLeuLe 242
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
669 GCACATTCATTTCATCGAGCGCTGCGCGGGAATAAACCGTGTGACCA 718
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
242 uArgThrLysTyr.....ProGlnGlyAlaGluLys..... 252
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
719 TCATTATCAAGAGGTGATTGCTATCGGACGCTTTGTCGTAACAGCGCGT 768
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
253 .....GlnLeuIleTyrAlaIle..... 258
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
769 CTGAATACCGAGCGCGTGGTTCGCTTGGCGCGCTGCAATCAACAACC 818
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
259 .....ThrGlyArgMetValProArgGlyGlyLeuProMetAspValG 273
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

819 GCGCTCTTGGCT.....ACCGTTTGGTGGCAAGGTGTCTCAAC 859
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
273 yValValValGlnAsnValGlyThrCysValAlaValLysGluAlaValV 290
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
860 TTACCGCGCGGCAATTTGGTTGACGGGACACCGCGGTGATTTCCGGTTCG 909
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
290 aLaspGlyLysProLeuValGluArgGlyMetThrValSerGlyAspAla 306
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
910 GTATTGAAC.....GGTGGCATTCACCAAGG 935
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
307 ValLysAsnGlnLysAsnLeuIleValArgIleGlyThrProValLysAs 323
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
936 CGCGCATGATTATTGGGACGCTTACCAATCAGATTTCCGTTATCGAAG 985
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
323 pValIleAspTyrCysGlyGly.....IleAspG 333
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
986 AAGCGCGCAGCAAGAGCTTTCGGCTGGTTCGCGCGGACCGCGGACAAA 1035
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
333 luAsnThrGluArgValIleLeuGly.....GlyProMetMetGly... 346
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1036 TACTCCATCACGCGCACCACTCTCGGCCATTTCCTAAAAAACAAACTCTT 1085
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
347 IleSerIleThrAsn.....LeuAspIleProValMe 357
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1086 CAAGTTCACGACGCGCTCAACGCG.....GGCG 1114
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
357 tLysGlyThrSerGlyIleThrAlaPheLeuProLysLysSerArgProG 374
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1115 ACCGCGCATGTTACCGACGCTATGAGCGCGTAAATCGCGTTGGAC 1164
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
374 InLysProCysIleArgCysSerGluCysValGlnValCysProMetAsn 390
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1165 ATCTGCTGCTACCTTTCGCGCATTTAATCGTGGCGATACCGACAG 1214
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
391 LeuGlnProTyrLeuLeuTyrLeuSerThrLysArgLysTyrAspG 407
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1215 CGCCGAGCTTGGTTCGTTGGAATTCGAGGAGAGACCTCGCTTGT 1264
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
407 uAlaValGluAsnGlyLeuMetAspCysIleGlu.....CysGlySerC 422
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1265 GCAGCTTCGTCGCGCGGCAATACGAATAC 1296
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
422 yThrThrThrCysProSerLysIleGluHis 432
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

seq_name: pir2:D83208

seq_documentation_block:

probable ferredoxin PA3491 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: D83208
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
.; Lory, S.; Olson, M.V.
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Reference number: A82950; MUID:20437337
A:Accession: D83208
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-774 <STO>
A:Cross-references: GB:AE004770; GB:AE00491; NID:99949633; PIDN:AG06879.1; GSPDB:GN
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA3491

alignment_scores:
Quality: 132.00 Length: 492
Ratio: 0.600 Gaps: 24
Percent Similarity: 44.715 Percent Identity: 20.935

alignment_block:

US-09-303-518D-131 x D83208

Align seg 1/1 to: D83208 from: 1 to: 774

```
28 CTGCCCATCGCGGCGAGACGCGAGCAAGTCATTTATGAGCGCGCGCGCAT 77
|||||
23 LeuProIleGlnAlaProLeuAlaGlnArgTyrIleValPro..... 37
78 TACCGAAGTCGCGTTCCTGGCGAAGATATATCGCG.....ATGCGCC 121
|||||
38 .....LeuGly...GlnHisIleGlyAlaProAlaArgp 48
122 CTCGATGAAATCAAGAGTGAAGCGTCAAAAAAGCCGCAAGTCGTG 171
||
48 roCysValGluVal.....GlyGlnAlaValLeuLysGlyGlnThrIle 62
172 TTTGAAGCAAAAAGATCCGGCGCTAGTATTACTGCGCGCGCTTCAGG 221
||
63 AlaLeuProAspGlyThrValSerAlaAlaLeuHisAlaProThrSerGI 79
222 CAAATCGCGCGTATT.....CACCGTGGCGAAGAAAGCGCG 256
|||||
79 yThrValValAlaIleGlyAlaHisProTyrProHisAlaSerGlyLeup 96
257 TACTTCAGTCAGTCGTGATTGCCGTGAAGCAAGCGAGAA..... 297
|||||
96 roAlaProAlaIleValIleAlaSerAspGlyLeuGluArgTrpThrGlu 112
298 .....ATCGAGTTCGAAGCTACGTACTGAAGCGCTGGCGCAA 335
|||||
113 LeuHisProCysProAspPheArgAlaGluSerProLeuAlaLeuLeu.. 128
336 ATTGAGCAGCAAAAAGTGC CGCG.....AACCTGATTCAAT 373
|||||
129 .....GluArgIleArgAlaAlaGlyIleGlyLeuGlyAla 142
374 CAGGCTTATGACATCGG.....CTTGCACCGCTGCTTCAAGCAAAATC 417
|||||
142 laGlyPheProThrAlaAlaLysLeuAlaAlaArgProAlaGluLysIle 158
418 CTGCCGTAGATGCCGAGCGCTTCGCCATCTTCGTCAATGATGGAGAC 467
|||||
159 .....HisThrLeuValValAsnGlyAlaGluCys 168
468 CAATCCGCTGCTCGGACCGCTACGTCATCAATAAAGACCGCGCAAG 517
|||||
168 sGluProTyrIleSerAlaAspLeuLeuMetArgGluArgAlaThrG 185
518 ACTTCAAAAGCGGCTGTGTGATTGAGCGCGCTGACCGACGAAATC 567
|||||
185 lvalLeuGlyGlyIleAspIleLeuValGlnIle..... 196
568 CATGTGTGTAAGCAGCAGCGCGACGCTGCGCTGTAATGCTGCCAA 617
|||||
197 ...LeuCys.....ProGluGluValLeuValGI 205
618 TATCGAACAACATGAA.....TTTGGCGGCGCG.... 645
|||||
205 yIleGluAspAspLysProGluAlaIleAlaAlaLeuGlyAlaAlaLeuG 222
646 .....CATCTCGCGCGC.....CATCTCGCGCGC 657
222 lYGluArgProTyrArgIleValAlaLeuProThrArgTyrProSerGly 238
658 .....TTGAGTGGCAGCAGCATTCATTCAT 683
239 GlyGluArgGlnLeuIleGlnLeuLeuThrGlyArgGluValPro...Al 254
684 CGAGCCAGTCGCGCGCAATAAAGCGGTGGACCATCAATTATCAAGACG 733
|||||
254 aspGlyLeuProAlaAspIleGlyIleLeuCysGlnAsnValGlyThrL 271
734 TGATTGCTATCGACGCTTTGTTGTAACAGCGCGTCTGAATACCGAGCGC 783
```

```
271 euAlaAlaValHisAspAlaValValLeuGlyArgProLeuIleSerArg 287
|||||
784 GTGTTGCTTGGCGCGCTGCAAGTCAACAACCGCGCTCTTGGGTAC 833
|||||
288 IleThrThrLeuAlaGlyGlyAlaLeuGluArgProMetAsnValGluAl 304
834 CTTTTGGTGGCGAAGGTGCTCAACTTACCGCG..... 867
|||||
304 aLeuIleGlyThrProValHisGluLeuLeuAlaPheAlaGlyLeuAlaG 321
868 ..GCCGAATTGGTACGCGACGACCGCTGATTTCGGTTCGGTATTG 915
|||||
321 luGlyArgLeu.....GluArgValLeuMetGlyGly..... 331
916 AACGTCGATTCACAAAGCGCGCATATTATTGGGACGCTACCACAA 965
331 ..... 331
966 TCAGATTTCGTTATCGAAGAAGCGCGCAGCAAGAGCTGTTCCGCTGG 1015
|||||
332 .....ProMetMetGlyPheA 337
1016 TTGCGCGCGACGCGACAAATCTCCATCACGCGCACCTCTCGGCCAT 1065
|||||
337 laLeuProAsp..... 340
1066 TTCTTAAATAAACAATCTTCAAGTTACGACAGCGCTCAACGCGCGCA 1115
|||||
341 ...LeuSerValProLeuIleLysThrCysAsnCysLeuLeuAlaGlyAs 356
1116 CCGCGCATGCTACCG.....ATCG 1135
356 pAlaThrGluLeuProGluProValProAlaMetProCysIleArgCysG 373
1136 GCATTATGAGCGCGTAATGCGCTTGGACATCTGCTTACCTTGTGTTG 1185
|||||
373 LyAspCysAlaGlnValCysProValSerLeuLeuProGln...GlnLeu 388
1186 CGGATTTAATCTCGCGGATACCGACGCGCGAGGCTTTGGTGTGCTT 1235
|||||
389 HisPheAlaLeuGlyAsp...GluHisGluGlnLeuLeuAlaHisAs 404
1236 GGAATTGGAGAAAGACCTCGCTTTGTGACAGCTTCGCTGCGCGGCA 1285
|||||
404 nLeuPheAspCysIleGluCysGlyAlaCysAlaTyrValCysProSerS 421
1286 AATACGAATACGCGCGCTGTTGCGC 1311
421 erIle.....ProLeuValGln 426
```

seq_name: pir2:A84943

seq_documentation_block:

membrane protein [imported] - Buchnera sp. (strain APS)

C:Species: Buchnera sp.

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001

C:Accession: A84943

R:Shigenobu, S.; Watanabe, H.; Hattori, M.; Sakaki, Y.; Ishikawa, H.

Nature 407, 81-86, 2000

A:Title: Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp

A:Reference number: A84930; MUID:20445173

A:Accession: A84943

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-473 <STO>

A:Cross-references: GB:AP000398; GSPDB:GN00144

A:Experimental source: strain APS

C:Genetics:

A:Gene: rnfC

alignment_scores:

828 GCGTACCGTTTGGTGGCGAAGGTGTCTCAACTTA 862
 266 uSerGlyAsnPhetrrpValargileglyThrProIleLysTyrPheLeuT 283
 863 CCGCCGCGGAATTGGTTGACGCGGACACCGCGTGAATTCGGTTCGGTA 912
 283 hrSerAsnLysLeuLysGlnSer.....PheIleAlaSerValTyr 296
 913 TTGAACGGTGGATTGCACAAAGCGCGCATGATTATTGGGAGCGTACCA 962
 297 LeuGlyglyProPheMetGlyLysLysI 306
 963 CAATCAGATTTCGGTTATCGAAGAAGCGCGCAGCAAGAGCTGTTCGGT 1012
 306 eAsnAsnLeuAsn..... 310
 1013 GGGTTCGCGCGCAGCGGACAAATCTCATCAGCGCCACCTCTCGGC 1062
 311HisSerIleLeuLysLysThrAsnSer 319
 1063 CATTTCCTAAACAACTCTTCAAGTTCACGACACCGCGTCAACGCGG 1112
 320 IlePheIleThrHisLysLysGluLysAsnGluSerIle.....Se 333
 1113 CGACCGCGCATGTCACCGATCGGCACCTTATGACGCGGTAAATCGCGTGG 1162
 333 rGluLysThrCysIleArgCysGlyTyrCysSerTyrValCysProVala 350
 1163 ACATCCCTGCCTACCTGCTT.....TTGCGCGATTAAATCGTCGCG 1203
 350 snLeuLeuProGlnGlnLeuTyrTrpTyrIleLysAsnLysAsnHisVal 366
 1204 CATACCGACAGCGCGCAGCGTTCGTTGGTTCGTTGGAATTCGACGAGAAGA 1253
 367 GlnThrLysLysHisTyrValLeuAspCysIleGlu..... 378
 1254 CTGCTGTTGTCAGCTTCGTCGCGCGGCAATACGAATACGACGCGCCG 1303
 379 .CysLysAlaCysGluLysValCysPro.....SerTyrIleProL 392
 1304 TGTTCGCGAAAGTGTCTGAA 1323
 392 euValLysTyrPheIleGln 398

seq_name: pir2:A40670
 seq_documentation_block:
 nuclear envelope protein POM 121 - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 05-Nov-1999
 C:Accession: A40670
 R:Hallberg, E.; Wozniak, R.W.; Blobel, G.
 J. Cell Biol. 122, 513-521, 1993
 A:Title: An integral membrane protein of the pore membrane domain of the nuclear enve
 A:Reference number: A40670; MUID:93328754
 A:Accession: A40670
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-1199 <HAL>
 A:Cross-references: GB:221513; NID:g9396746; PIDN:CAA79725.1; PID:g9396747
 F:803-807,845-849,956-960,1010-1014,1047-1051,1076-1080/Region: pentapeptide motif (X

alignment_scores:
 Quality: 129.50 Length: 478
 Ratio: 0.576 Gaps: 20
 Percent Similarity: 47.071 Percent Identity: 22.385
 alignment_block:
 US-09-303-518d-131/rev x A40670 ..
 Align seg 1/1 to: A40670 from: 1 to: 1199

127 ATGAAATCAAGCAAGGTGAAGCGCTCAAAAGGCCAA...GTGCTGTT 173
 20 leuArgValLysIleAsnGluLysValLeuArgGlyGlnProLeuIlePh 36
 174 TGAACAAAAAAGATTCGCGCGGTAGTATTACTCGCGCGCTTCAGCA 223
 36 eSerAspPheAsn.....ValProValHisAlaProThrSerGlyL 51
 224 AAATCGCCCTATTACCGTGGCGGAAAGCGGTACTTCAGTCAGTCGTG 273
 51 euIleGluAsnIleCysPheAsnSerAspSerIleLysLysAsnIleLys 67
 274 ATTCCGCTTGAAGGC.....AACGACGA 296
 68 IleValIleSerProAspTyrIleuAspGlnTrpIleArgLeuAsnProI 84
 297 AATCGAGTTCGACGCTAGCTACCTGACGCGTGGCAAAATGACGAGCG 346
 84 eLysAspTyrLysTyrAlaProGluLysLeuIleLysIle..... 98
 347 AAAAAGTGGCGCGCAACCTGATTCATCAATCAGGCTTATGAGCTCGCTCGC 396
 99IleHisGlnSerGlyVal...ValGlyLeuGly 108
 397 ACCCGTCCGTTTC.....ACCAAAATCCCTGCGGTAGATCCGCGAGCC 437
 109 GlyGlyGlnPheProSerSerLysLysIleIlePheSerIleAsnArgAl 125
 438 GTTCGCGCATCTCGTCAATCGATGACACCAATCGCTGGCTGCGGACG 487
 125 aHisThrLeuIleValAsnAlaValGluSerGluProTyrIleThrSera 142
 488 CTACGCTCATCATCAAGAACGCGCGAAGACATTCAAACGCGCGCTGTTG 537
 142 spAsnCysLeuIleTyrAsnHisIleSerGluIleLeuIleGlyCysLys 158
 538 GTATTGAGCGCGCTGACCGACGTAATCCATGCTGTGTAACGACGACG 587
 159 IleIleCysTrpIleThrLysIleLysThrValLeuIleAlaIleGlnG 175
 588 CGCAGAGTCCGCTCT.....GAAATGCTG 613
 175 uAspAsnIleGlnSerIleSerLysIleGlnHisLeuIleLysAsnLys 192
 614 CCAATATCGAACAACATGAATTTGGCGCGCGCATCTCGCGCG..... 657
 192 erLeuPheLysIleCysIleIleLysLysTyrProAlaGlySerSer 208
 658TTGAGTGGCAGCACATTT.....CA 677
 209 LysValLeuValLysSerLeuThrGlyLysGluValProHisGlyLysH 225
 678 TTTTCATCGAGCAGTTCGCGCGCAATAAACCGTGTGGACCATCAATATC 727
 225 sSerIleAsp.....TyrLeuIlePheAsnVala 237
 728 AAGACGTGATTCTATCGGAGCTTGTTCGTAACAGCGCGCTGCAATACC 777
 237 laThrIlePheSerIleLysArgAlaIleLysAsnGlyLysProLeuThr 253
 778 GAGCGCGTGTGCTTGGCGGCTGCAAGTCAACAAACGCGCGCTT 827
 254 GluArgValValThrLeu.....MetSerAspLysAsnLeuLe 266

```

801 LysProValPheGlyPheGlyValThrThrAlaAlaSerThrAlaSerTh 817
      :::||| ::: ::::||||||| ||| |||
492 CGTAGGTCGGCAGCC...ACGGGATGGTGTCCATCGCATTCAGCAAGA 446
      |:::||||: ||| :||| ||| :|||
817 rIleAlaSerThrSerGlnSerIleLeuPheGlyAlaProProValT 834
      :||| :||| :||| :||| :|||
445 TGGCGAAGCGCTCGGCA...TCTACGCAGCGGATTTTCTGAACGCGG 399
      ||||:||||| :||| :|||:||||: |||
834 hrAlaSerSerAlaProAlaLeuAlaSerIlePheGlnPheGly... 849
      :||| :||| :||| :||| :|||
398 GTGCGAAGCGCAGTCCATAGCCCTGATTGAATCAGGTTGCGGCGC 349
      ||||| LysPro...
850 ..... LysPro... 851
348 TTCGTGCTCAATTTTCCAGCGGCTTCAGGTCAGTACGCTTCGAAC 299
      ||| ||||:||||: ||||| ||||:||||:
852 ....LeuAlaProAlaAlaSerValAlaGlyThrSerPheSerGln 867
      :||| :||| :||| :||| :|||
298 TTTTCGTCGTTGCCTTCAACGCGCAATCAGGACTGACTGAAGTAC 249
      :||| :||| :||| :||| :|||
867 euAlaSerSerAlaGlnThrAla..... 874
248 TCGCCACGGTGAATAGCGGCGATTTTCCTGAAGCCGCGCAGTAA 199
      |||||
875 ..... 878
198 TAGCCCGCGGATTC.....TTTTGTGCTTCAAACAGCAGCTT 164
      :||| :||| :||| :||| :|||
878 rSerGlyGlyPheSerGlyPheGlyThrLeuThrThrSerThrSer 895
      :||| :||| :||| :||| :|||
163 GGCCTTTTTCAGCGGCTTCACCTCTCTGATTTTCATCGAGGCGCG 114
      |||| :||| :||| :||| :|||
895 laProAlaThrThrSerGlnProThrLeuThrPheSerAsnThrVal 911
      :||| :||| :||| :||| :|||
113 CGCACATATCTTCGCCAAGCAACGCGACTTCGGTAATGCCGGCGC 64
      ||||:||||: ||| :||| :|||
912 ProThrPheAsnIleProPheSerAlaSerAlaLysProAlaLeu 928
      :||| :||| :||| :||| :|||
63 ATAATAGACTTCTCGGTCGCTCGCGCGATGGGC 30
      : :|||:||||: ||||:||||:
928 rTyrProGlyAlaAsnSerGlnProThrPheGly 939
      : :|||:||||: ||||:||||:
seq_name: pir2:S69703
seq_documentation_block:
HKR1 protein precursor - yeast (Saccharomyces cerevisiae)
N/Alternate names: protein YDR420W
C/Species: Saccharomyces cerevisiae
C/Date: 22-Aug-1996 #sequence_revision 06-Sep-1996 #text_change
C/Accession: S69703; A53382
R/Dietrich, F.S.
submitted to the EMBL Data Library, August 1995
A/description: The sequence of S. cerevisiae lambda 3641 and cosmid
A/reference number: S69555
A/Accession: S69703
A/Molecule type: DNA
A/Residues: 1-1802 <DIE>
A/Cross-references: EMBL:U33007; NID:g927685; PTDN:AAB64857.1; P:
R/Kasahara, S.; Yamada, H.; Mio, T.; Shiratori, Y.; Miyamoto, C.
J. Bacteriol 176, 1488-1499, 1994
A/title: Cloning of the Saccharomyces cerevisiae gene whose over
A/reference number: A53382; MUID:94156857
A/Accession: A53382
A/Molecule type: DNA
A/Residues: 1-581, 'A', 583-593, 'A', 595-1802 <KAS>
A/Cross-references: EMBL:S69101; NID:g545659; PTDN:AAB30051.1; P:
A/Experimental source: YNNA295
A/Note: sequence extracted from NCBI backbone (NCBIN:144410, NCBB
C/Genetics:
A/Gene: SGD:HKR1
A/Cross-references: SGD:S0002828; MIPS:YDR420W
A/Map position: 4R

```


A: Residues: 1-519 <SCH>
A: Cross-references: EMBL:X72888; NID:g435523; P1DN:CAA51399.1; PID:g435525
C: Genetics:
A: Gene: rnfC
C: Superfamily: unassigned ferredoxin 2[4Fe-4S]-related proteins; ferredoxin 2[4Fe-4S] h
C: Keywords: iron-sulfur protein
F: 374-438/Domain: ferredoxin 2[4Fe-4S] homology <FER>

```

alignment_scores:
  Quality: 126.50
  Ratio: 0.586
  Percent Similarity: 47.285
  Length: 457
  Gaps: 18
  Percent Identity: 18.381

alignment_block:
  US-09-303-518D-131 x S39893
  ..
  Align seg 1/1 to: S39893 from: 1 to: 519

```

70 CCGCGCATACCGAAGTGGCGTTGCTTGGCGAAGAATATGTCCGCATCGC 119
||||| : : : : ||| : : : : : : : : : :
42 ProAlaLeuIleArgLeuProLeu.....GlnGlnHisIleGlyAlaGI 56

120 CCCCTCGATGAATCAAGGAAGTGAAAGCGCTCAAAAAGAGGCCAAGTCG 169
:
56 uAlaGluProIleValLysArgAspLeuValLeuLysGlyGlnLeuI 73
:
170 TGTGTTGAAGACACAAGAATCCGGCGCTAGTATTACTCGCGCGCTTCA 219
:
73 leAlaLysAlaArgGlyProLeuSerAlaAsnIleHisAlaProThrSer 89
:
220 GGC AAAATCGCGCGTATT.....CACCGTGGCGAAAA 251
||| : : : : ||| : : : :
90 GlyArgValIleAlaValGlyHisPheValAlaProHisAlaSerGlyLe 106
:
252 CGCGGTACTTCAGTCAGTCGTGATGTCGGTTGAAGGCAACGACGAA... 297
||| :
106 uProVal...ProthrIleThrIleArgProAspGlyGluAspLysTrpG 122
:
298ATCGAGTTCAACGCTACGTACCTGAAGCGCTG 330
:
122 lyProHisLeuProArgLeuArgProGluAsnAlaAlaProGluGluIle 138
:
331 GC AAAATTGAGCAGCGAAAAAGTCGCCGCAACCTGATTCAATCAGGCTT 380
||| :
139 Ala.....AlaGlnValAlaAlaAglyIleValGlyMetGlyGI 152
:
381 ATGGACT.....CGCTTCGCACCCTCCGTTCCAGCAAATCCCTG 421
||| :
152 yAlaThrPheProSerAlaValLysLeuAsnLeuArgAlaLysTyAspL 169
:
422 CCGTAGATCGCGAGCGCTTCGCCATCTTCGTCAATCGATGGACACCAAT 471
:
169 euThr.....ThrLeuIleIleAsnGlyAlaGluCysGlu 180
:
472 CGCGTGGCTGCCAGCCCTACGGTCATCATCAAAAGAAGCCCGAAGACTT 521
||| :
181 ProTyLeuThrCysAspArgLeuMetArgGluargAlaGluGluII 197
:
522 CAACCGCGCCTGTTGGTATTGACCGCTGACCGAACGTAATAATCCATG 571
: ||| :
197 eAlaAspGlyIleGlyIleMetaIaArgAlaLeuGlyValLys...GlnV 213
||| :
572 TGTGTAAGCAGCAGCGAGCGTCCGCTCTCAAAATGCTGCC..... 615
||| :
213 alPheValAlaIleGluSerAsnLysProGlnAlaIleGluAlaMetThr 229
:
616AATATCGAAACACATGAATTTGGCGG 641
:
230 ArgTyAsnArgAlaLeuGlyTyThrPheLysIleHisValValProth 246
:
642 CCGCGCATCTCGCGGCTTGAGTGGCAGCGACATTCATTATTCATGAGCCAG 691
: : : : | | | :

246 rGlnTyrPro.....MetGlySerGluLysHisLeuValLysMetI 260
692 TCGGCGCGAATAAACCC.....GTG 711
:: : :: :: :: ::
260 l eThrGlnGluThrProAlaArgAlaLeuThrAlaAspLeuGlyVal 276
:: : ||| :: : ||| ::
712 TGGACCATCAATATCAAGACGCTGATTCGTCGACGTTTGTTTCGTAAC 761
:: : ||| :: : ||| ::
277 ValValHisAsnIleAlaThrAlaHisAlaValHisLeuAlaValArgTy 293
762 AGGCCGCTCTGAATACCGACGCGTGTGCCTTGGCGGCGCCTCAACTCA 811
|||| :: ||| :: ||| :: ||| ::
293 rGlyGluProLeuIleAlaArgThrValThrValSerGlyHisGlyIleA 310
812 ACAACCGCGCCTCTGGGTACCGTACCGTTTTGGGTGCGAAGGTGTCACAATT 861
:: : ||| |||| :: :: :: :: ||| :: ||| :: ||| ::
310 rgArgProAlaAsnLeuArgValLeuIleGlyThrProValSerGlyIleA 326
862 ACGGCC.....GGCGAATTGGTTACCGCGGACACACCGCTGATTTCCGG 905
||| ||| :: :: :: :: :: :: :: :: :: ::
327 IleAlaHisCysGlyGlyPheThrGluGluProAspArgLeuLeuGlu 343
||| :: :: ||| |||
343 yGlyProMetMetGlyMetProIleGln..... 352
956 GCTACCACAAATCAGATTTCCGTTATCGAAGAGCCGACGAAAGACTG 1005
:: : :: :: :: ||| :: :: :: ::
353AsnProValProValValLysGly..... 361
1006 TTCGGGTGGTTTGGCCGCGACGACAAAATACTCCATCAGCGCACCCAC 1055
361 361
1056 TCTCGGCCATTCTCAA AAAACAACACTCTTCAAGTTCCAGCACGCCGTCA 1105
||| :: :: :: :: :: :: :: :: ::
362ThrAsnGlyIleLeuAlaLeuThrAlaAlaGluT 373
1106 ACGCGCGGACGCGCCCATG.....GTACCGATCGGCACCTTATGAGCGC 1149
:: : ||| :: :: :: :: :: :: ::
373 hrProGluAlaLysThrMetProCysIleArgCysGlyArgCysValGln 389
1150 GTAATGCCGTGGACATCCTCCCTACCTTGTCTTTGCGCGATTAATACGT 1199
||| :: :: ||| ||| :: :: ::
390 GlyCysProValGlyLeuThrProPheGluLeuAsnAlaArgIleHisAl 406
1200 CGSCCATACCGACGCGCGCAGGCTTTGGGT.....TGC TTGGAAT 1240
: |||| :: :: :: :: :: :: ||||
406 aGlyAspLeuGlyAlaAlaLysValGlyLeuMetAspCysLeuAla. 422
1241 TGGACGAAGAAGACCTCGTTGTCAGCTTCGCTGCCCGGGCAATAAC 1290
:: : |||| :: :: |||| :: :: ::
423CysGlyCysCysSerTyrAsnCysProAlaAsnLeu 434
1291 GAATACGCCCGCTGTTCGC 1311
|||| :: :: :: ::
435ProLeuValGln 438

seq_name: pir2:T43481

seq_documentation_block:
probable mucin DKFZp434C196.1 - human (fragment)
N:Alternate names: protein DKFZp434B0635.1

C/Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 20-Apr-2000
C/Accession: Homo sapiens (man)
C/Accession: T43481, T43459, T17264
R/Kocher, K.; Beyer, A.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, December 1999
A/Reference number: Z22514
A/Accession: T43481
A/Molecule type: mRNA
A/Residues: 1-580 <AAA>
A/Cross-references: EMBL:AL133561; NID:q6599133; PIDD:CAR63715.1; PID:q6599133

A: Experimental source: adult testis; clone DKFp434C196
 R: Poustka, A.; Wellenreuther, R.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
 submitted to the Protein Sequence Database, October 1999
 A: Reference number: Z21540
 A: Accession: T34549
 A: Molecule type: mRNA
 A: Residues: 262-580 <POU1>
 A: Cross-references: EMBL:AL122069; NID: g6102864; PIDN: CAB59245.2; PID: g7018420
 A: Experimental source: adult testis; clone DKFp434B0635
 R: Poustka, A.; Klein, M.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
 submitted to the Protein Sequence Database, September 1999
 A: Reference number: Z18723
 A: Accession: T17264
 A: Molecule type: mRNA
 A: Residues: 262-580 <POU2>
 A: Cross-references: EMBL:AL117481; NID: g5911958; PIDN: CAB55954.1; PID: g5911959
 A: Experimental source: adult testis; clone DKFp434B061
 C: Genetics:
 A: Note: DKFp434C196.1; DKFp434B0635.1; DKFp434B061.1

alignment_scores:
 Quality: 121.50 Length: 530
 Ratio: 0.557 Gaps: 23
 Percent Similarity: 41.132 Percent Identity: 23.019

alignment_block:
 US-09-303-518d-131/rev x T43481 ..

Align seg 1/1 to: T43481 from: 1 to: 580

```

1336 CCTCTCAATGGTTTCAGCACTTCGCGAACAGCGGCGGCTATTCGTAT 1287
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
29 ProSerArgAlaSerProThrArgMetProSerArgAlaSerLeuLysMe 45
1286 TTGCGCGGCGAGCAGAGTGCACAAAG.....CGAGGTCTTCTTC 1246
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
45 tThrProPheArgAlaSerLeuThrLysMetGluSerThrAlaLeuLeuA 62
1245 GTCAATTCAC..... 1235
|| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
62 rgThrLeuProArgAlaSerLeuMetArgThrProThrArgAlaSerLeu 78
1234 .....AGCAA 1230
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
79 MetArgThrProProArgAlaSerProThrArgLysProProArgAlaSe 95
1229 CCCAAG...CTGCGGCGTGTGGTATCCCGACGATTAATCGGCAC 1183
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
95 rProArgThrProSerArgAlaSerProThrArgArgLeuProArgAlas 112
1182 AAGCAAGTAGGAGGATGTCCACGCAATACGCGCTCATAGTGCAGA 1133
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
112 erPro.....MetGlySerProHis.....ArgAlaSerProMetArg 124
1132 TCGGTACCATGGCGCGTGCAGCGGTGACGGCTGTCGTAACCTGAAG 1083
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
125 ThrProPro.....ArgAlaSerProThr..... 132
1082 AGTTTGTTTTAGGAATGCCGAGAGTGGTGGCGCTGATGAGTATTT 1033
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
133 .....GlyThrProSerThrAlaSerProThrGlyT 143
1032 GTCCGGCTGGCGCAACCGACGCAACAGCTCTTTGCTGGCGCTTCTT 983
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
143 hrProSerSerAlaSerProThrGlyThrProPro..... 154
982 CGATAACGAAATCTGATGTGGTAGCGTCCCAATTAATCATGGCCCT 933
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
155 .....ArgAlaSerProThrGlyThrProProArgAlaThr 169
932 TGTGCAATCGACCGTTCATA.....CCGAACCGGAATCA.... 896
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

169 gSerProSerThrAlaSerLeuThrArgThrProSerArgAlaSerLeuT 186
895 .....CGCGGTGT 887
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
186 hrArgTPrProProArgAlaSerProThrArgThrProProArgGluSer 202
886 CCGCGTCAACCAATCCCGCGGTAAAGTTGAGACACCTTC.....GCACC 841
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
203 ProArgMetSerHisArgAlaSerProThrArgThrProProArgAlaSe 219
840 CAAAACGGTACGCAAGAGCGCGTGTGACCTGCGAGCCGCCCAAGG 791
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
219 rProThrArgArgProProArgAlaSerProThrArgThrProProArg 236
790 CAACACCGCTCGGTATTTCAGACGGCTGTTCAGAAACACCTCCGATA 741
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
236 luSerLeuArgThrSerHisArgAlaSerProThrArgMetProProArg 252
740 GCAATCAGCTTGTGATAATTTGATGTCACACAGGTTTTATTTCGGCGC 691
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
253 AlaSerProThr.....ArgArgProProArgAlaSerProTh 265
690 TGGCTCGATGAATGATGTCGTCACCAAGCGCGAGGATGCGGG. 642
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
265 rGlySerProPro.....ArgAlaSerProMetThrProProArgAlaSerP 281
641 .....CGCCCAAAATTCATGT 627
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
281 roArgThrProProArgAlaSerProThrThrProSerArgAlaSer 297
626 GTTCGATATTGGCAGCATTTTCAGACGGCAGCTCTCGCTGCTGCTTT 577
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
298 LeuThrArgThrProSerTrpAlaSerProThrThrProSerArgAl 314
576 ACACACATGGATTTTACGTCGTCAGCGGCTCAATACCAACAGCGCCG 527
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
314 aserLeuMetLysMetGluSerThrValSerIleThrArgThrProA 331
526 GTTGAAG.....TCT 516
|| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
331 rgAlaSerProThrGlyThrProSerArgAlaSerProThrGlyThrPro 347
515 TCGCGCGCTTCTTCATGATACCGTAGGTCGCGCAGCCAGGATGGT 466
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
348 SerArgAlaSerLeu.....ThrGlySerProSerArgAlaSerLeuTh 362
465 GTCCATCGATTGACGAAGATGGCGACGCTCGCATCTACGCGAGGA 416
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
362 rGlyThrProSerArgAlaSerLeuIleGlyThrProSerArgAlaSerL 379
415 TTTTCTGACAGCGGCGGTGGAGCGGAGCGGAGCCATACGCTGATGATC 366
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
379 euIleGlyThrProSerArgAlaSerLeuThrGlyThrProProArgAla 395
365 AGTTTGGCGCGCACTTTTCGCTGCTCAATTT..... 333
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
396 SerLeuThrGlyThrSerSerThrAlaSerLeuThrArgThrProSerAr 412
332 .GCCAGCGCTCAGGTACGTTCGAACTCGAATTCGTTTCGTTGCCTT 284
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
412 gAlaSerLeuThrArgThrGlnSerSerSerSerLeuThrArgThrProS 429
283 CAACGGCAATCAGCTGACTGACTGAGTACGCTTTTCGCGCAGGTGAATA 234
|| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
429 erMetAlaSerLeu.....ThrArgThrProProArg..... 439
233 GCGGCGATTTTG.....CCTGAAGCGCGCAGTAAATACCTACGCC 193
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
440 AlaSerLeuThrArgThrProProArgAlaSerLeuThrArgThrProPr 456
192 CGAATCTTTTGTCTTCAACAGCAGCTTGGCCTTTTTCAGCGCTTCA 143
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
456 oArgAlaSerLeuThrArgThrProProArgAlaSerLeuThrArgThrP 473

```



```
seq_name: pir2:F85769
382 rLeu 383

seq_documentation_block:
probable membrane protein 22636 [imported] - Escherichia coli (strain O157:H7, substrain
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Dec-2001
C:Accession: F85769
R:Perina, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimantanta, E.; Potamouisis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: F85769
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-740 <STO>
A:Cross-references: GB:AE005174; NID:gl2515618; PIDN:AAG56618.1; GSPDB:GN00145; UWGP:226
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
C:Superfamily: unassigned ferredoxin 2[4Fe-4S]-related proteins; ferredoxin 2[4Fe-4S] hc

alignment_scores:
Quality: 121.00 Length: 319
Ratio: 0.742 Gaps: 14
Percent Similarity: 51.097 Percent Identity: 22.571

alignment_block:
US-09-303-518D-131 x F85769
Align seg 1/1 to: F85769 from: 1 to: 740
100 GAAGAATATGTGGCATGCGCCCTCGATGAAATCAAGGAAGTGAAGC 149
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
48 LysGlnHisIleGlyAlaGluGlyGluLeuCysValSerValGlyAspLy 64
150 CGTCAAAAGGCGCAAGTGTGTTGAAGACAAAGATCCGGCGCTAG 199
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
64 sValLeuArgGlyGlnProLeuThrArgGlyArgGlyLysMetLeuPro 81
200 TATTACTGCGCGCTTCAGGAAATCGCGCTATT..... 237
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
81 alHis....AlaProThrSerGlyThrValThrAlaIleAlaProHisSer 96
238 .....CACCGTGGCGAAAGCGCGTACTTCAGTCAGTCGTGATGCCGT 281
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
97 ThrAlaHisProSerAlaLeuAlaGluLeu...SerValIleIleAspAl 112
282 TGAAGGCAACGACGAATCGAGTTCGAACGCTAGCTACCT...GAAGCGC 328
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
112 aspGlyGluasp.....CysTrpIleProArgAspGlyT 124
329 TGGCAAAATTCAGCGCAAAAGTGGCGCCCAACCTGATTCAA..... 372
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
124 rPalaAspTyrArgSerArgArg...ArgGluGluLeuIleGluArgile 139
373 .....TCAGGCTTATGGACTGGCT 392
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
140 HisGlnPheGlyValAlaGlyLeuGlyAlaGlyAlaGlyPheProThrGlyVa 156
393 TCGCACCCTCGCTTCAGCAAAATCCCTCGCGTAGATCGCGAGCGGTTCG 442
156 l.....LysLeuGlnGlyGlyGlyAspLysIleGluT 167
443 CCATCTTCGTCAATCGCATGACCAACCAATCGCGTGGTGGCGACCCCTACG 492
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
167 hrLeuIleIleAsnAlaAlaGluCysGluProTyrIleThrAlaAspAsp 183
493 GTCATCATCAAGAAGCGCGGAAGACTTCAACGCGGCTGTGGTATT 542
```

```
184 ArgLeuMetGlnAspCysAlaAlaGlnValValGluGlyIleArgileLe 200
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
543 GAGCGCGCTGACCGAA..... 558
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
200 uAlaHisIleLeuGlnProArgGluIleLeuIleGlyIleGluAspAsnL 217
559 .....CGTAAATCCATGTGTGAAGCAGCAGCGCGCAGACGTCGCTCT 603
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
217 ysProGlnAlaIleSerMetLeuArgAlaValLeuAlaAsp..... 230
604 GAAATGCTGCCAATATGCAACACATGAATTTGGCGCGCGCGCATCTGCG 653
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
231 .....SerHisAspIleSerMetArgValIleProThrLysTyrProSe 245
654 CGGC.....TTGAGTGGCAGCAGCACATTTCATT 679
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
245 rGlyGlyAlaLysGlnLeuThrTyrIleLeuThrGlyLysGlnVal.... 260
680 TCATCGACCGCATCGCGCGCAATAAAACCGTGTGGACCATCAATATCAA 729
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
261 .....ProHisGlyGlyArgSerSerAspIleGlyValLeuMetGln 274
730 GACGTG.....ATTGCTATCGGACGTTTTCGTAAACAGCGCGCTCT 770
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
275 AsnValGlyThrAlaTyrAlaValLysArgAlaValIleAspGlyGluPr 291
771 GAATACCGAGCGCGTGTTCGTTGGCGCGCTGCAAGTCAACAAACCGC 820
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
291 olleThrGluArgValValThrLeuThrGlyGluAlaIleAlaArgProG 308
821 GCCTCTTCGTACCGTTCGTTGGTGGAGGTGTCTCAACTT.....ACC 864
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
308 lYAsnValTrpAlaArgLeuGlyThrProValArgHisLeuLeuAsnAsp 324
865 GCGCGCGATTTGTTGACGCGACACCGCGTATTTCGTTCCGTTCCGTTAT 914
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
325 AlaGlyPheCysProSerAlaAspGlnMetValIleMetGlyGlyProLe 341
915 GAACGCT 921
341 uMetGly 343
seq_name: pir2:B90921

seq_documentation_block:
probable membrane protein [imported] - Escherichia coli (strain O157:H7, substrain RI
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 14-Dec-2001
C:Accession: B90921
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and 9
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: B90921
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-772 <HAY>
A:Cross-references: GB:BA000007; PIDN:BAR35761.1; PID:gl3361805; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RIMD 050952
C:Genetics:
C:Superfamily: unassigned ferredoxin 2[4Fe-4S]-related proteins; ferredoxin 2[4Fe-4S]

alignment_scores:
Quality: 121.00 Length: 319
Ratio: 0.742 Gaps: 14
Percent Similarity: 51.097 Percent Identity: 22.571

alignment_block:
US-09-303-518D-131 x B90921
```

Align seg 1/1 to: B90921 from: 1 to: 772

```
100 GAAGAAATATGTCGGCATCGCCCTCGATGAAATCAAGGAGGTGAAC 149
    : : : : : : : : : : : : : : : : : : : : : : : : :
48 LysGlnHisIleGlyAlaGluGlyLeuCysValSerValGlyAsp 64
    : : : : : : : : : : : : : : : : : : : : : : : : :
150 CGTCAAAAGGCCAAGTCCTGTTGAAGACAAAGAAATCCGGCGTAG 199
    : : : : : : : : : : : : : : : : : : : : : : : : :
64 sValLeuArgGlyGlnProLeuThrArgGlyLysMetLeuProV 81
    : : : : : : : : : : : : : : : : : : : : : : : : :
200 TATTTACTCGCCGCTTCAGCAAAATCGCCGCTATT..... 237
    : : : : : : : : : : : : : : : : : : : : : : : : :
81 alHis...AlaProThrSerGlyThrValThrAlaIleAlaProHis 96
    : : : : : : : : : : : : : : : : : : : : : : : : :
238 .....CACCGTGGCGAAAGCGCTACTTCAGTCAGTCGATCGCGT 281
    : : : : : : : : : : : : : : : : : : : : : : : : :
97 ThrAlaHisProSerAlaLeuAlaGluLeu...SerValIleIleAsp 112
    : : : : : : : : : : : : : : : : : : : : : : : : :
282 TGAGGCAACGACGAAATCGAGTTCGAACGCTAGCTACT...GAAGCG 328
    : : : : : : : : : : : : : : : : : : : : : : : : :
112 aAspGlyGluAsp.....CysTrpIleProArgAspGlyT 124
    : : : : : : : : : : : : : : : : : : : : : : : : :
329 TGGCAAAATTCAGCAGCGAAAGTGGCCGCCCACTGATCAA..... 372
    : : : : : : : : : : : : : : : : : : : : : : : : :
124 rPalaAspTyrArgSerArg...ArgGluGluLeuIleGluArg 139
    : : : : : : : : : : : : : : : : : : : : : : : : :
373 .....TCAGGCTTATGGACTGCGCT 392
    : : : : : : : : : : : : : : : : : : : : : : : : :
140 HisGlnPheGlyValAlaGlyLeuGlyAlaGlyPheProThrGly 156
    : : : : : : : : : : : : : : : : : : : : : : : : :
393 TCGCACCGCTCGTTCAGCAAAATCCCTGCGTAGATGCCGAGCGCTCG 442
    : : : : : : : : : : : : : : : : : : : : : : : : :
156 l.....LysLeuGlnGlyGlyAspLysIleGluT 167
    : : : : : : : : : : : : : : : : : : : : : : : : :
443 CCATCTTCGTCAATGCGATGCACACCAATCCGCTGGCTGCCGCCCTACG 492
    : : : : : : : : : : : : : : : : : : : : : : : : :
167 hrLeuIleIleAsnAlaGluCysGluProTyrIleThrAlaAsp 183
    : : : : : : : : : : : : : : : : : : : : : : : : :
493 GTCATCATCAAGAACGCGGAGACTTCAAACGCGCGCTGTGGTATT 542
    : : : : : : : : : : : : : : : : : : : : : : : : :
184 ArgLeuMetGlnAspCysAlaAlaGlnValValGluGlyIleArg 200
    : : : : : : : : : : : : : : : : : : : : : : : : :
543 GAGCCGCTGACCGAA..... 558
    : : : : : : : : : : : : : : : : : : : : : : : : :
200 uAlaHisIleLeuGlnProArgGluIleLeuIleGlyIleGluAsn 217
    : : : : : : : : : : : : : : : : : : : : : : : : :
559 .....CGTAAATCCATGTGTGTAAGCAGCAGCGCAGCGTCCGCTCT 603
    : : : : : : : : : : : : : : : : : : : : : : : : :
217 ysProGlnAlaIleSerMetLeuArgAlaValLeuAlaAsp..... 230
    : : : : : : : : : : : : : : : : : : : : : : : : :
604 GAAATGCTGCCAATATCGAAACACATGAATTTGGCGCGCCGATCTGC 653
    : : : : : : : : : : : : : : : : : : : : : : : : :
231 .....SerHisAspIleSerMetArgValIleProThrLysTyrPro 245
    : : : : : : : : : : : : : : : : : : : : : : : : :
654 CGCG.....TTGAGTGGCAGCCACATTCATT 679
    : : : : : : : : : : : : : : : : : : : : : : : : :
245 rGlyGlyAlaLysGlnLeuThrTyrIleLeuThrGlyLysGlnVal... 260
    : : : : : : : : : : : : : : : : : : : : : : : : :
680 TCATCGAGCGATCGCGCGGAATAAAACCGTGTGGACCATCAATTATCA 729
    : : : : : : : : : : : : : : : : : : : : : : : : :
261 .....ProHisGlyGlyArgSerAspIleGlyValLeuMetGln 274
    : : : : : : : : : : : : : : : : : : : : : : : : :
730 GAGCTG.....ATTGCTATCGAGCGTTTTCGTAAACAGCGCTCT 770
    : : : : : : : : : : : : : : : : : : : : : : : : :
275 AsnValGlyThrAlaTyrAlaValLysArgAlaValIleAspGlyGlu 291
    : : : : : : : : : : : : : : : : : : : : : : : : :
771 GAATACGAGCGCTGTTGCTTGGCGCGCTCGCAAGTCAACAACCCG 820
    : : : : : : : : : : : : : : : : : : : : : : : : :
291 oileThrGluArgValValThrLeuThrGlyGluAlaIleAlaArgPro 308
    : : : : : : : : : : : : : : : : : : : : : : : : :
821 GCCTCTGGGTACCGTTTGGGTGGCAAGGTGTCTCAACTT.....ACC 864
    : : : : : : : : : : : : : : : : : : : : : : : : :
```

308 lyAsnValTrpAlaArgLeuGlyThrProValArgHisLeuLeuAsnAsp 324

865 GCCGGCGAATTTGGTTGACGGGCAACACCGCGTGATTTCCGGTTCGGTATT 914

325 AlaGlyPheCysProSerAlaAspGlnMetValIleMetGlyGlyProLe 341

915 GAACGGT 921

341 uMetGly 343

seq_name: pir2:I52257

seq_documentation_block:

episialin - mouse

C:Species: Mus musculus (house mouse)

C:Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 05-Nov-1999

C:Accession: I52257; I65210

R:Vos, H.L.; De Vries, Y.; Hilkens, J.

Biochem. Biophys. Res. Commun. 181, 121-130, 1991

A:Title: The mouse episialin (MucI) gene and its promoter. Rapid evolution of the rep

A:Reference number: I52257; MUID:92068178

A:Accession: I52257

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-631 <RES>

A:Cross-references: GB:M77226; NID:g199835; PIDN:AAA39754.1; PID:g199837

A:Accession: I65210

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-631 <RES>

A:Cross-references: GB:M84683; NID:g199842; PIDN:AAA39756.1; PID:g199843

C:Genetics:

A:Gene: MucI

A:introns: 20/1; 454/3; 472/2; 517/1; 557/3; 607/3

alignment_scores:

Quality: 119.00 Length: 456

Ratio: 0.569 Gaps: 22

Percent Similarity: 45.833 Percent Identity: 24.123

alignment_block:

US-09-303-518D-131/rev x I52257 ..

Align seg 1/1 to: I52257 from: 1 to: 631

1337 TCCTCTCAATGGTTCCAGCACTTTGCGCATTCGCAACACGCGCGCGTATTCGTA 1288

38 SerSerSerLeuAlaSerThrThr..... 45

1287 TTTGCGCGGCGACAGCACTGCACAAAGCGAGGTCTTCTTCGTCCTCAATT 1238

46ThrProValHisSer.....SerAsnSerAspProA 56

1237 CCAAGCAACCCAAAGCCTGCGCGTATCGCGATCGCGACGATTAATTCG 1188

56 laThrArgProGlyAspSerThrSer...SerProValGlnSerSer 71

1187 CGCAAAACGAGGTAGCAGGATGTCACCGCATTCACCGCTACATAAGT 1138

72 ThrSerSerProAlaThrArgAlaProGluAspSerThrThrAlaVa 88

1137 GCGCATCGGTACCATGCGCGCTGCGCGCTGACGCGCTGCTCGTGA 1088

88 lIeuSerGlyThr.....SerSerProAlaThrThrAlaProValAsn 103

1087 TCAGAGATTGTTTTTAGGAAATGGCGAGAGTGGTGGCGGTGATGGAG 1038

103 erAlaSer.....Ser.....Ser 106

1037 TATTGTGCGCGTGGCGGCAACCCAGCCGACAGCTCTTTGCTGCGGCC 988

107 ProValAlaHisGlyAspThrSerSerProAlaThrSerLeuSerLys 123

```
987 TTCTTCGATAACGAAATCTGATTTGTTGGTACGTCCTCCCAAAATATCATCGG 938
||||| : : : : : |||
123 pSerAsnSerProVal.....ValHisSerGlyThrSerA 137
||||| : : : : : |||
937 CGCCTGTGCAATCGCAGCGTTCAATACCGAACCGGAAATACACGCGGTG 888
||||| : : : : : |||
137 laProAlaThrThrAlaProValasp..... 145
887 TC CGCGCTCAACCAATTCGCGCGGTAGTTGAGACACCTTCGACCCAA 838
||||| : : : : : |||
146 .....SerThrSerProValValHisGlyThrSerSerProAla 160
837 AACGGTACCAAGACGCGGTTTGTGACTTCAGCGCGCCCAAGGCA 788
||||| : : : : : |||
160 aThr.....SerProGlyaspSerThrSerProAspHisSers 175
787 CCACGCGCTCGGTATTACAGACGCGCTTTACGAACAAACGTCGATAGCA 738
||||| : : : : : |||
175 erThrSer.....ProAlaThr....ArgAlaProGluasp 186
737 ATCAGCTCTTGATAATTGATGTCACACGGTTTATTTCGCGCGGACTGG 688
||||| : : : : : |||
187 SerThrSer.....ThrAlaValLeuSerGlyThrSe 197
687 CTCGATGAATGAATGTCGTCGCCACTCAAGCGCGGAGTGGCGCGC 638
197 rSer.....ProAlaThrThrAlaProv 205
637 CAATTCATGTTGTCATATTGGCA..... 612
205 alAspSerThrSerProValAlaHisAspThrSerSerProAla 221
611 ...GCATTTTCAGAC.....GGCAC 595
222 ThrSerLeuSerGluAspSerAlaSerSerProValAlaHisGlyThr 238
594 GTCTGCGCTGCTGCTTTACACATGGATTTTACGTTTCGTCAGCGCGC 545
||||| : : : : : |||
238 rSerSerProAlaThr.....SerProLeuArgA 248
544 TCATACCAACAGCGCGCTTGAAGTCTTCGCGGCTCTTTGATGATG 495
248 spSerThrSerProValHisSerAlaSerleGlnAsnIleLys 264
494 ACCGTAGGTCGCGCAGCAGC.....GGATTGGTGTCCATCGC 457
265 ThrThrSerAspLeuAlaSerThrProAspHisAsnGlyThrSerValTh 281
456 ATTGACGAAGATGCGGAACGCGCTCGGCA..... 429
281 rThrThrSerAlaLeuGlySerAlaThrSerProAspHisSerGlyT 298
428 .....TCTACGCGCAGGATTTGCTGAACGCGGCGTG 396
298 hrSerThrThrThrAsnSerSerGluSerValLeuAlaThrThrProVal 314
395 CGAAGCGCAGTCCATAAGCCTGATTGAATCAGGTTGCGCGGCACATTTC 346
315 TyrSerSerMet.....ProPheSe 321
345 GCTGCTCAATTTGCCAGCGCTTACGATAGTACGCTTCG.....AATCGA 299
321 rThrThrLysValThrSerGlySerAlaIleProAspHisAsnGlys 338
298 TTTGCTGCTGCTTCAAGCGCAATCAGACTGACTGAAGTACGCGCTTT 249
338 erSerValLeuProThrSerSerValLeuGlySer..... 349
248 TCGCACGGTGAATAGCGCGGATTTTCGTCGAACGCGCGGAGTAATAC 199
350 .....AlaThrSerLeuValTyrAsnThrSerAlaIleAlaTh 362
```

```
seq_name: pir2:T05352
seq documentation_block:
hypothetical protein F8B4.120 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 23-Jul-1999
A:Accession: T05352
R:Bevan, M.; Terry, N.; Ardiles, W.; Buysshaert, C.; Dasseville, R.; De Clerck, R.;
ewes, H.W.; Mayer, K.F.X.; Schueller, C.
submitted to the Protein Sequence Database, February 1999
A:Reference number: Z15409
A:Accession: T05352
A:Molecule type: DNA
A:Residues: 1-857 <BEV>
A:Cross-references: EMBL:AL034567
A:Experimental source: cultivar Columbia; BAC clone F8B4
C:Genetics:
A:Map position: 4
A:introns: 26/3; 45/1; 74/3; 83/1; 122/2; 165/1; 270/2; 307/1; 731/2; 754/2
A:Note: F8B4.120
C:Superfamily: cyclophilin homology
F:6-162/Domain: cyclophilin homology <CYP>
```

```
alignment_scores:
Quality: 118.00 Length: 438
Ratio: 0.621 Gaps: 21
Percent Similarity: 43.379 Percent Identity: 24.429
alignment_block:
US-09-303-518D-131 x T05352 ..
Align seg 1/1 to: T05352 from: 1 to: 857
```

```
111 CGGCATGCGCCCTCGATGAAATCAAGAAAGGTGAACCCGTCAA..... 155
||||| : : : : : |||
448 ArgGluSerProGlySerGluGlyArgHisValArgArgSerPr 464
156 .....AAAGGCCAAGTCTGTTTGAAGACAAAAGAAAT 189
464 oThrLysSerValSerArgSerProValArgValLysLysGluArgAspI 481
190 CGGGCGGTAGTATTACTGCGCGGCTTCAGCAAAATCGCGCTATTTC 239
481 leSerArgSerPro.....SerLysSerLeuSerArgSer 492
240 CCGTGGCGAAAGCGCGTACTTCAGTCTAGTCTGAT...TGCCGTTGAG 286
493 ProLeuArgSerProLysArgValIleSerArgSerProValArgGly 509
287 GCACGACGAAATCGAGTTTCGAACG...CTAGTACTGAACGCGTGGCA 333
509 gIleAlaArgSerProSerArgSerProValArgSerAlaSerArgGlys 526
334 AAATTGAGCAG..... 344
526 erLeuGlyArgGlyProLeuArgArgSerSerArgSerProSerArg 542
345 .....CGAAAGTCGCGCGCAACCTGATTCAATCAGCGCTTATGA 385
543 SerProValArgSerSer.ArgArgSerLeuSerArgSerProIleGlnL 559
```

100 GAAGATATGTCGGCATGCGCCCTCGATGAATAACAAGAAAGGTGAAGC
LysGlnHisIleLeuAlaGluGlyLeuLeucysValSerValGlyAspLeu
150 CGTCAAAAAGGCCAAGTGCTTTTGAAGACAAAAAANAATCCGGCGCTAG
sValLeuArgGlyGlnProLeuThrArgLyArgGlyLyMetLeuProv
200 TATTACTCGCGCGCTTCAGCAAAATCGCGGTATT
alHis...AlaprothrSerGlyThrValThrAlaIleAlaProHisSer
282 TGAAGGCACGACGAAATCGAGTTCGAACGCTAGTACCTGAA
aspGlyGluasp.....CystTrpIleProArgAspClyT
325CGCTGGCAAAATGACGACGGAANAAGTCGCGCGCAACCTGATT
rpAlaaspTyrargThrArgSerArgGluGluLeuIleGluArgIleHis
370 CAATCAGGCTATGACCTGCCTTCGACCGCTCCGCTTCAGC
GlnPheGlyVal...AlaGlyLeuGlyClyAlaGlyPheProThrClyVa
412 AAATCCCTCGCTAGATGCGGAGCGGTTTCGCCATCTTCGTCGAATGCGA
llysLeuGlnGlyGlyAspylIleGluThrLeuIleLeuAlaA
156 llysLeuGlnGlyGlyAspylIleGluThrLeuIleLeuAlaA

alignment_scores:
Quality: 116.00 Length: 441
Ratio: 0.518 Gaps: 19
Percent Similarity: 50.794 Percent Identity: 22.222
alignment_block:
US-09-303-518D-131/rev x 147141 ..
Align seg 1/1 to: 147141 from: 1 to: 528
1343 CAGCCTTCCTCTCAATG.....GTTCCAGCAGCTTGGCGCAACAG 1303
149 GlnProSerSerSerProProIleSerThrVal..... 162
1302 CGGCGCGTATTTCGTATTTCGCCGGGAGAGAGCTGCACAAAGCGAGT 1253
163SerValGlnProSerSerSer.....SerAlaPro 174
1252 CTCTTCGTCCTCAATTCAGCAACCCAAAGCCTGCGCGTGTGCGTATCG 1203
174 hrThrSerAlaThrSerValGlnProSerSer.....SerSerSer 187
1202 CGCAGCATTAATTCGCGCAAAAGCAAGGTAGCGAGGTGTCACACGCAT 1153
188 ProProIleSerSerThrValSerValGlnThrSerSerSerSerVa 204
1152 T.....ACGCGCTCATAAGTGCCGATGCGGTACCGTGGCGGGT 1115
204 lProThrThrSerThrSerValGlnPro.....SerSerSerSer 219
1114 CGCGCGGTGACGCGTCTGCTGAACCTGAAGAGTGTGTTTTTAGGAAA 1065
219 erValProThrThrSerAlaThrSerValArgSerSerSerSerSer 235
1064 TGGCGGAGAGTGTGCGCGTATGAGTATTTGTCGGCTCGCGCGCAAC 1015
236 ThrPro.....IleProSerThrThrSerVa 244
1014 CCAGCGCAACAGCTCTTGTGCGGCGCT...TCCTCGATACCGAATCT 968
244 lGlnProSerSerSerSerAlaProThrThrSerAlaThrSerVal. 260
967 GATTGTGTAGCTCCCAATAATCATGCGCGCTTGTGCAATCGCACCG 918
261GlnProSerSerSerSerThrProIleProSerThr 273
917 TTCAATACCGAAGCAATCACGCGTGTGTCGCGCTCAACCAATTCGCC 868
274 ThrSerValGlnPro.....SerSerSerSerSerAlaPr 285
867 GCGGTAAAGTTGAGACACCTTCGCACCCCAACAGGTACGCAAGGCGCG 818
285 ofThrThrSerAlaThrSerValGlnProSerSer..... 296
817 GTTCTGTCAGCTGCGCGCGCAAGGCAACGCGCTCGGTATTCAGA 768
297SerSerSerProIleSerSerThrIleSerValGlnPro 310
767 CGGCGCTGTACGACAAACGTCGATACGATACGCTCTTGATATGAT 718
311 SerSerSerSerSerProThrThrSerThrSer.....ValGl 325
717 GTCACACAGGTTTATTTCGCGCGCGCTGCGATGAAATCAATGTGCG 668
325 nProSerSerSerGlySerAlaProThrThrSer.....Ala 338
667 TCCCACTCAAGCGCGCAGGATGCGCGCGCAAAATTCATGTGTTTCGATA 618
338 hrSerValGlnProSerSerSerSerProIleSerSerThrIle 354
617 TTGGCAGCATTTTCAGACGCGCAGCTGCTGCGCTGCTGTTTACACATG 568
355 SerValGlnProSerSerSerSerProThrThrSerThrSe 371

567 GATTTCACGTTTCGTCAGCGGCTCAATACCAACAGCGCGTGTGAAGT 518
 ::: :
 371 rValGlnProSerSerG 378
 517 CTTCCGCGGCTTCTTTGATGATGACC.....GTAGGTCGCGACGCCAGC 474
 : : : : :
 378 lySerAlaProThrThrSerAlaThrSerValGlnProSerSerSer 394
 : : : : :
 473 GGATTGGTGTCCATCGCATTCGACGAAGATGCGGACGCTCGGATCTAC 424
 : : : : :
 395 SerValProThrThrSerAlaThrSerValArgSerSerSerSe 411
 : : : : :
 423 GCGAGGATTTGCTGAACGACGGTCCGA.....AGCGCAG 386
 : : : : :
 411 rThrProleProThrThrThrSerValGlnProSerSerSerSe 428
 : : : : :
 385 TCCATAACCCCTGATGATCAGGTTCGGCGCACATTTTCGCTGCTCAAT 336
 : : : : :
 428 alProThrThrSerAlaThrSerValGlnThrSerSerSerThr 444
 : : : : :
 335 TTTGCCAGCGCTTCAGGTACGTCGAACTCGATTTTCGTCGTTGCC 286
 : : : : :
 445 ProleProSerThrThrSerValGlnProSerSerSerSerAlaPr 461
 : : : : :
 285 TTCACGCGCAATCAGACT...GACTGAAGTACGCGCTTTTCGCCACGGT 239
 : : : : :
 461 oThrThrSerAlaThrSerValGlnProSerSerSerPro... 476
 : : : : :
 238 GAATAGCGCGGATTTTCGCTGAACCGCGCAGTAAATACATACGCCGGA 189
 : : : : :
 477ProleSerSerThrThrSerValGlnPro... 486
 : : : : :
 188 TTTCTTTTGTCTTCAACAGCAGCTTCGCTTTTTCGCGCTTCACCTTC 139
 : : : : :
 487SerSerSerSerSerProThrThrThrThrThrSerVa 500
 : : : : :
 138 CTTGATTTTCATCAGCGGCGCATCGGACATATCTTCGCCAAGCAAG 89
 : : : : :
 500 lGlnProSerSerSerGlySerAlaProThrThrSerAlaThrSerValG 517
 : : : : :
 88 CGACTTCGGTAATGGCGGCGG 66
 : : : : :
 517 lNProSerSerSerSerPro 524

seq_name: pir2:T34513

seq_documentation_block:
 hypothetical protein ZK783.1 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 C:Accession: T34513
 R:Favella, A.; Vaudin, M.
 submitted to the EMBL Data Library, August 1994
 A:Description: The sequence of C. elegans cosmid ZK783.
 A:Reference number: Z21536
 C:Accession: T34513
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-3507 <FAV>
 A:Cross-references: EMBL:U13646; PIDN:AAC24418.1; GSPDB:GN00021; CESP:ZK783.1
 A:Experimental source: strain Bristol N2; clone ZK783
 C:Genetics:
 A:Gene: CESP:ZK783.1
 A:Map position: 3
 A:Introns: 14/1; 48/2; 84/1; 196/3; 303/1; 381/1; 586/1; 605/1; 1175/3; 1207/1; 1409/2;
 3504/1

alignment_scores:

Quality:	115.00	Length:	415
Ratio:	0.948	Gaps:	20
Percent Similarity:	50.602	Percent Identity:	22.410

alignment_block:

US-09-303-518D-131/rev x T34513

Align seg 1/1 to: T34513 from: 1 to: 3507

1283 CCCGGCAGACGAGAGTGCAC.....AAAGCAGGTCTTCTTCGTCCAA 1240
 : : : : :
 1945 ProAsnArgLysIleGluIleAspGluGluAsnSerSerSerAs 1961
 : : : : :
 1239 TTCACGAACCAACCAAGCCTGCGCGTGTGCGGTATCCCGACGATTAAT 1190
 : : : : :
 1961 nSerGlyGlnGluLys.....ProThrThrLysG 1971
 : : : : :
 1189 CGGCGCAACCAAGGTAGGAGGATTCACACGGCATTACGCGCTCATAA 1140
 : : : : :
 1971 lyIleValSerSerThrSerAlaThrSerSerGluSerThrThrAlaGlu 1987
 : : : : :
 1139 GTCCGATCGGTACCATGCGCGGTGCGCGCTTCACGCTGCTCGTGA 1090
 : : : : :
 1988 ProHisValThrThr.....SerIleSerSerThrThrSe 1999
 : : : : :
 1089 CTTGAACAGTTTGTCTTTTGGAAATGCGGAGAGTGTGCGCGTCATGG 1040
 : : : : :
 1999 rThrLysAspMetThrSerSerLysSerProGluAsnValThrMetSerS 2016
 : : : : :
 1039 AGTATTTGTCGGCTCGCGCGCAACCCGCGGTAAGTTGACACACC 990
 : : : : :
 2016 erGluSerProGluValSerThrSerSerLysSerThrThrAlaSer 2032
 : : : : :
 989 CTTCTTCGATAACGAAATCTGATTGTGTGAGCGTCCCAATAATCATG 940
 : : : : :
 2033 GluThrThrValSerSer.....ThrProSerGluSerSe 2044
 : : : : :
 939 C.....GCGCCTGTGCAATCGCACCGTTCAATACCGCAACGGAAA 899
 : : : : :
 2044 rSerSerGluAlaProLeuThrSerSerProAlaThrThrThrGluValI 2061
 : : : : :
 898 TCACGGGGTGTCCGCGTCAACCAATTCGCGCGGTAAAGTTGACACACC 849
 : : : : :
 2061 leThrGluSerSerValLysSerThrThrProLysGlu.....GluSer 2075
 : : : : :
 848 TTCGCACCCAAACGCTAGCAAGAGCGCGGTGTTGACTTCGACGCGCC 799
 : : : : :
 2076 SerSerGluIleThrValLys.....LeuSerSerLysSe 2087
 : : : : :
 798 GCCCAAGCAACACGCGCTCGGTATTCAGACGCGCTGTAGC...AACA 752
 : : : : :
 2087 rProGluValThrGluSerSerValLysSerSerProSerThrProSerT 2104
 : : : : :
 751 AACGTCCGATACCAATCATGCTTCGATAATGATGTCCTCCACACGGTTTA 702
 : : : : :
 2104 hrThrSerGluSerValThrSer.....ThrVal... 2113
 : : : : :
 701 TTCGCGCGCGACTGCGCTCGATGAATGATGTCGTCACCAAGCCGCGC 652
 : : : : :
 2113 2113
 651 AGGATCGCGCGCGCAAAATTCATGTTTCGATATTGCGCAGATTTTCAG 602
 : : : : :
 2114ProGluThrThrSerLysSerThrValLeuSerSerGluAlap 2127
 : : : : :
 601 ACGGCACGCTCT...GCGCCTGCTGCTTTACACACATGATGATTTTACGTCG 555
 : : : : :
 2127 roValThrSerThrSerProThrGluValHisThr..... 2138
 : : : : :
 554 GTCAGCGCGCTCAATACCAACAGCGCGGTGTTGAAGTCTTCGCGCGCTTC 505
 : : : : :
 2139SerSerGluThrLysProSerLeuSerAlaSerSer... 2150
 : : : : :
 504 TTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 473
 : : : : :
 2151ThrThrGlyAspThrAsnSerThrThrProSerThrSerS 2164

```
472 GATTGGTGTCCATCGATTGACGAGATGCGCAAGCGC...TCGCGCATCT 426
:::|||||::: |||||::: |||||::: |||||::: |||||:::
2164 erLeuAlaSerValLysSerThrSerAlaProGluGlyThrSerAlaSer 2180
:::|||||::: |||||::: |||||::: |||||::: |||||:::
425 ACGGAGGAGGATTTGCTGAACGAGCGGGTGCGAAGCGCAGTCCATAAGCC 376
:::|||||::: |||||::: |||||::: |||||::: |||||:::
2181 ValAlaProValLysLeuSerSer...LeuSerProAspValSerGlnPr 2196
:::|||||::: |||||::: |||||::: |||||::: |||||:::
375 TGATTGAATCAGGTTGCGCGCAGCTTTTCGCTGCTCAATTTTGGCAGCG 326
:::|||||::: |||||::: |||||::: |||||::: |||||:::
2196 oSer.....ThrLysThrPheAsp.....AlaThrG 2205
:::|||||::: |||||::: |||||::: |||||::: |||||:::
325 CTTTCAGTACGTAGCTTCGACTCGATTTCGTTGCTTCACAGCGCA 276
:::|||||::: |||||::: |||||::: |||||::: |||||:::
2205 luSerSerValGlnAlaSerGluThrSerSerGlyThrSerValLys 2221
:::|||||::: |||||::: |||||::: |||||::: |||||:::
275 ATCACCAGCTGACTGAAGTACGGCGCTTTTCGCCACGCTGAATAGCGCGCAT 226
:::|||||::: |||||::: |||||::: |||||::: |||||:::
2222 SerThrSerGluProGluSerHisValThrLysLeuSerIleThrSerSe 2238
:::|||||::: |||||::: |||||::: |||||::: |||||:::
225 TTTGCTGAAGCGCGCGCAGTAAATACTACGCCCGGATTTCTTTTGTCTTT 176
:::|||||::: |||||::: |||||::: |||||::: |||||:::
2238 rAsnProSerSerValProValThrSerPro.....LysS 2251
:::|||||::: |||||::: |||||::: |||||::: |||||:::
175 CAACAGCAGCTTGGCGCTTTTTCAGCGCTTCACCTTCCTTGATTTCATC 126
:::|||||::: |||||::: |||||::: |||||::: |||||:::
2251 erThrProThrValProGluSerThrGluGlnProThrSerThrPro 2267
:::|||||::: |||||::: |||||::: |||||::: |||||:::
125 GAGGGCGCATGCGGACATATTCTTCCCAAGCAAGCGGACTTCG 81
:::|||||::: |||||::: |||||::: |||||::: |||||:::
2268 SerGlyGlnSerLeuThr.....ProMetAsnSerAsnSer 2279
:::|||||::: |||||::: |||||::: |||||::: |||||:::
seq_name: pir2.A45294

seq_documentation_block:
Balbani ring 2.1 - midge (Chironomus tentans) (fragment)
C:Species: Chironomus tentans
C>Date: 25-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jul-2000
C:Accession: A45294
R:Wieslander, L.; Paulsson, G.
Proc. Natl. Acad. Sci. U.S.A. 89, 4578-4582, 1992
A:Title: Sequence organization of the Balbani ring 2.1 gene in Chironomus tentans.
A:Reference number: A45294; MUID:92262483
A:Accession: A45294
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-749 <WIES>
A:Cross-references: GB:M89909; NID:g156605; PIDN:AAA28264.1; PID:g156606
A>Note: sequence extracted from NCBI backbone (NCBIN:102269, NCBIN:102271, NCBIN:102273).
C:Superfamily: unassigned Balbani ring proteins
C:Keywords: tandem repeat

alignment_scores:
Quality: 114.50 Length: 476
Ratio: 0.500 Gaps: 26
Percent Similarity: 48.109 Percent Identity: 23.109

alignment_block:
US-09-303-518D-131 x A45294 ..
Align seg 1/1 to: A45294 from: 1 to: 749
12 CAATAAGGTTCTAATCTGCCATCGCGGCAGAC..CGGAGCAAGTCATT 60
:::|||||::: |||||::: |||||::: |||||::: |||||:::
317 GluLysArgSerLysSerGlySerArgProGluLysArgSerLysSerG1 333
:::|||||::: |||||::: |||||::: |||||::: |||||:::
61 TATGACGGCGCGCCATTACCGAAGTCCGCTGCTTGGCGAAGAATATG 110
:::|||||::: |||||::: |||||::: |||||::: |||||:::
333 ySerArgProGluLysArgSerLysSerGlySerArgProGluLysArgS 350
:::|||||::: |||||::: |||||::: |||||::: |||||:::
111 CGGATGGCGCCCTCGATGATAAATCAAGGAAGTGAAGCGCTCAAAAAG 160
:::|||||::: |||||::: |||||::: |||||::: |||||:::
```


962 pValGlyLysThrThrLeuIleGlnAsnPheGluAspLeuAsnAspProT 979

299 TCGAGTTTCGAACGCTACGTACCTGAAGCGCTGC AAAATTGAGCAGCGAA 348

979 yrProValAlaAlaArgPheLeuAspAlaTy rProGluAlaSerThrGln 995

349 AAAGTCGCCCGCAACCTGATTCAATCAGGCTTATGGACTGCGCTTCGCAC 398

996 AspLeu.....AsnThrGlnAspAlaGlnPhePheTy rSerLeuCyse 1010

399 CCGTCGCTTTCAGCAAA.....ATCCCTCCGCTAGAT..... 429

1010 rAsnProPheGlnLysProValProPheIleProAlaIleAspThrP 1027

430GCCGAGCCGCTTCGCC 444

1027 heGluPheTy rPheLysLysAspSerLeuTrpGlnSerGluAspLeuAla 1043

445 ATCTCGTCAAT.....GCGATGGACACCAATCCGCT 476

1044 AlaValValGlyLysPValGlyArgValAlaIleLeuGlnGlyProMe 1060

477 GCGTCGCCGACCTACGGTCATCATCAAAAGACGCCGCAAGACTTCAAAC 526

1060 tAlaAlaLysHisSerThrLysValAsnGluProAlaLysGlu..... 1074

527 GCGCGCTGTTGGTATTGACCCCGCTGACCGCAACGTAAATTC....CATGTG 573

1075LeuLeuAspGlyIleAsnGluThrHisIleGlnHisPhe 1087

574 TGTAAAGCA.....GCAGGCGCAGAC.....GTGCGCTGCAAAATGC 611

1088 IleLysLysPheTy rAlaGlyAspGluLysLysIleProLle..... 1101

612 TGCCAATATCGAAACACATGATTTGGCGGCCCGCATCTCGCGGTTGA 661

1102ValGluTy rPheGlyGlyValProProValAsnValS 1114

662 GT.....GGCAGCGCAC 672

1114 erHisLysSerLeuGluSerValSerValThrGluGluAlaGlySerLys 1130

673 ATTCATTTCATCGACCGACTCGCGCGCAATAAACCGTG..... 711

1131 ValTy rLysLeuProGluIleGlySerAsnSerAlaLeuProSerLysLy 1147

711 711

1147 sLeuTrpPheGluLeuLeuAlaGlyProGluTy rThrTrpPheArgAlaI 1164

712TGGACCATCAATTATCAAGAC 732

1164 lePheThrThrGlnArgValAlaLysGlyTrpLysLeuGluHisAsnPro 1180

733 GTGATTGCTATC.....GGACGTTTGTTCGTACACAGG 764

1181 ValArgArgIlePheAlaProArgTy rGlyGlnArgAlaValValLysGI 1197

765 CCGTCTGAATACCGAGCGCGTGTTCCTTGGCGGCCCTGCAAGTCAACA 814

1197 yLysAspAsn...AspThrValValGluLeuTy rGluThrGlnSerGly. 1212

815 AACCGCGCTCTTCGCTACCGTTTGGGTGCGAAGGTGTCCTCAACTTACC 864

1213AsnTy rValLeuAlaAlaArgLeuSer..... 1221

865 GCCGCGCAATTGGTTGACCGCGCAACCGCTGATTTCGGTTTCGGTATT 914

1222TyrAspGlyGluThrIleValValSerMetPheGluAs 1234

915 GAACGGTCGATTGCACAAGCCGCAT...GATTATTTGGACGCTAC. 960

1234 nArgAsnAlaLeuLysGlyValHisLeuAspPheLeuPheLysTy rG 1251

C; Superfamily: yeast fatty-acyl-CoA synthase beta chain
C; Keywords: acetyl-CoA; acyltransferase; carbon-oxygen lyase; coenzyme A; fatty acid bio

217 TCAGGCAAAATCGCGCTATTACACGCTGGCGGAAAAGCGGTACTTCTCAGTC 266
 ThrGlyLysLeu,.....MetArgIleGluGluArgPheThrClnAs 962
 267 AGTC.....GTGATTCGGCTTGAAGGC.....AACGACGAA 298
 :||| :||| :||| :||| :|||

Mon Jul 1 09:25:39 2002

us-09-303-518d-131.rpr

```

961 .....CACATCAGATTTCCTTATCGAAGAGCGCGCAGC 996
1251 luProSerAlaGlyThrSerProValSerGluLeuAspGlyArgAsn 1267
997 AAAGAGCTG.....TTCGGCTGGGTT.....GCCCGCA 1025
1268 AspArgIleLysHisPheThrAlaLeuThrPheGlyGluGluProTy 1284
1026 CGCGCAAAATCTCATCACCGCACCACTCTCGCCATTTCCCTAAAA. 1074
1284 rProGluAsnAlaSerIleThrAspThrPheThrGlyProGluValThrV 1301
1075 .....AACAACTCTTCAAGTTACGACAGCCGTCACCGCGCGCAG 1116
1301 alThrGlyAsnMetIleGluAspPheCysArgThrValGlyAsnHisAsn 1317
1117 CGCGCATGGTACCGATCGCGCTATGAGCGCTAATCGCGTTGGAC... 1164
1318 GluAlaThrThrLysArgAlaIleArgLysArgMetAlaProMetAspPh 1334
1165 .....ATCCTGCCTTACCT 1177
1334 eAlaIleValValGlyTrpGlnAlaIleThrLysAlaIlePheProLysA 1351
1178 TGCTTTTGGCGGATTTA.....ATCGTCGGCGATACCGACGCGCGCAG 1221
1351 laIleAspGlyAspLeuLeuArgLeuValHisLeuSerAsnSerPheArg 1367
1222 GCCTTGGGTGCTTGAATTTGGACGAGAGAGAC 1254
1368 MetValGlySerHisSerLeuMetGluGlyAsp 1378

seq_name: pir2:T39207

seq_documentation_block:
fatty acid synthase, subunit beta - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
C:Accession: T39207
R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Ransperger, U.; Pohl, T.
submitted to the EMBL Data Library, September 1999
A:Reference number: 221835
A:Accession: T39207
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-2073 <WOO>
A:Cross-references: EMBL:AL110469; PIDN:CAB54157.1; GSPDB:GN00066; SPDB:SPAC926.09c
A:Experimental source: strain 972h-; cosmid c926
C:Genetics:
A:Gene: SPDB:SPAC926.09c
A:Map position: 1
C:Superfamily: yeast fatty-acyl-CoA synthase beta chain

alignment_scores:
Quality: 114.50 Length: 461
Ratio: 0.556 Gaps: 25
Percent Similarity: 44.685 Percent Identity: 21.475

alignment_block:
US-09-303-518d-131 x T39207 ..
Align seg 1/1 to: T39207 from: 1 to: 2073

217 TCAGCAAAATCGCGCTATTCACCGTGGCGAAGAGCGGCTTTCAGTC 266
948 ThrGlyLysLeu.....MetArgIleGluGluArgPheThrGlnAs 962
267 AGTC.....GTGATTGGCTTGAAGC.....AACGACGAAA 298
962 pValGlyLysThrThrLeuIleGluAsnPheGluAspLeuAsnAspProt 979

```

```

299 TCAGATTGCAACGCTAGCTACCTGCAAGCGCTGGCAAAATTTAGCAGCGAA 348
979 yrProValAlaAlaArgPheLeuAspAlaThrProGluAlaSerThrGln 995
349 AAAGTCGCGCGCAACCTGATTCAATCAGGCTTATGGACTGCGCTTCGCAC 398
996 AspLeu.....AsnThrGlnAspAlaGlnPhePheThrSerLeuCysSe 1010
399 CCGTCCGCTTTCAGCAA.....ATCCCTGCGCTAGAT..... 429
1010 rAsnProPheGlnLysProValProPheIleProAlaIleAspThrP 1027
430 .....GCCGAGCCGCTTCGCC 444
1027 heGluPheThrPheLysLysAspSerLeuTrpGlnSerGluAspLeuAla 1043
445 ATCTTCGTCTCAAT.....CGCATGGACACCAATCGCT 476
1044 AlaValValGlyGluAspValGlyArgValAlaIleLeuGlnGlyProMe 1060
477 GGCTGCGGACCTACGCTCATCATCAAGAGCGCGCAAGACTTCAAC 526
1060 tAlaAlaLysHisSerThrLysValAsnGluProAlaLysGlu..... 1074
527 CGCGCTTGTGTTATGAGCGCTGACCGCAACGTAATC....CATGTG 573
1075 .....LeuLeuAspGlyIleAsnGluThrHisIleGlnHisPhe 1087
574 TGTAAGCA.....GCAGGCGCAGAC.....GTGCGCTCTGAAAATGC 611
1088 IleLysLysPheThrAlaGlyAspGluLysLysIleProIle..... 1101
612 TGCCAATATCAAAACACATGAAATTTGGCGCGCCCATCTGCGCGCTGA 661
1102 .....ValGluThrPheGlyGlyValProValAsnValS 1114
662 GT.....GCACGCGAC 672
1114 erHisLysSerLeuSerValSerValThrGluGluAlaGlySerLys 1130
673 ATTCTATTCATCAGCAGCCAGTCGGCGCAATAAAACCGTG..... 711
1131 ValTyLysLeuProGluIleGlySerAsnSerAlaLeuProSerLysLy 1147
711 ..... 711
1147 sleuTrpPheGluLeuLeuAlaGlyProGluTrpThrTrpPheArgAlaI 1164
712 .....TGGACCATCAATATTCAGAC 732
1164 lePheThrThrGlnArgValAlaLysGlyTrpLysLeuGluHisAsnPro 1180
733 GTGATTGCTATC.....GGACGTTTGTTCGTACAGG 764
1181 ValArgArgIlePheAlaProArgTyGlyGlnArgAlaValValLysGl 1197
765 CGCTCTGAATACCGCGCTGTTGCTTGGCGCGCTGCAAGTCAACA 814
1197 yLysAspAsn...AspThrValValGluLeuLeuTyGlnSerGly. 1212
815 AACCGCGCTCTTGGCTACCGCTTTTGGTGGCGAAGGTGCTCAACTTACC 864
1213 .....AsnTyThrValLeuAlaAlaArgLeuSer..... 1221
865 GCGCGCAATTTGTTACCGCGCAACCGCTGATTTCGGTTCGGTATT 914
1222 .....TyrAspGlyGluThrIleValValSerMetPheGluAs 1234
915 GRACGTCGATTCGCACAGGCGCGCAT...GATTATTGGGACGCGCTAC. 960
1234 nArgAsnAlaLeuLysGluValHisLeuAspPheLeuPheLysTyG 1251
961 .....CACAAATCAGATTTCCTTATCGAAGAGCGCGCAGC 996

```



```

311 rgArgArgGlyAspGlnHisArg..... 318
948 TTGGGACGCTACCAATCAGATTTCCTTATCGAAGAGCGCCGACGA 997
319 .....ProGlyGlnGluValArgValArgArgPro..... 329
998 AAGAGCTGTTCGGCTGGGTTGGCGCGCAGCGGACAAATACCTCCATCAGG 1047
330 .....AlaGlyProArgValVal..... 335
1048 CCACCACTCTCGGCATTTCTTAAACAACTCTTCAAGTTCCAGC 1097
336 .....SerArgSerAlaAlaArgAspAlaAlaArgGlnProAlaArg 349
1098 AGCCGTCACAGG.....CGGCGACCGCGCCATGCTACCGATCGGCA 1138
350 AspArgGluArgAlaValGlyArgArgProArg...GlyArgGluArgHis 365
1139 CTTATGAGCGGTATGCGGTGGACATCTGCCTACCTTCTGTTTGGGC 1188
365 s.....AlaAlaArgAlaSerGlyArgAlaGlyGluAlaHisGlyP 380
1189 GATTTAATCGTCGGGATACCGACAGCGCGCA 1220
380 roAspAspGlyArgArgCysAlaValArgAla 390

```

seq_name: pir2:T30635

seq_documentation_block:
hypothetical protein 33L - Molluscum contagiosum virus 1

N;Alternate names: MC033L

C;Species: Molluscum contagiosum virus 1

C;Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 21-Jul-2000

C;Accession: T30635

R;Senkevich, T.G.; Bugert, J.J.; Sisler, J.R.; Koonin, E.V.; Darai, G.; Moss, B.

A;Title: Genome sequence of a human tumorigenic poxvirus: Prediction of specific host re

A;Reference number: 220876; MUID:96325459

A;Accession: T30635

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-579 <SEN>

A;Cross-references: EMBL:U60315; NID:g1491943; PIDN:AAC55161.1; PID:g1491976

C;Genetics:

A;Note: MC033L

C;Superfamily: Molluscum contagiosum virus 1 hypothetical protein 33L

alignment_scores:

Quality	Length
112.00	473
Ratio: 0.563	Gaps: 23
Percent Similarity: 42.072	Percent Identity: 21.142

alignment_block:

US-09-303-518D-131 x T30635 ..

Align seg 1/1 to: T30635 from: 1 to: 579

```

115 ATGGCGCCCTCGATGAATCAAGAGGTGAAGCGCTCAAAAAA..... 159
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1 MetArgProHisValLeuIleThrLeuAlaThrCysAlaLeuArgAlaLe 17
160 .GGCAAGTCTCTTTGAACACAAAAAGATCCGGCGGTAGTATTACTG 208
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
17 uAlaGlnValIleAspGluHisGlnHisSerGluProProValSerThr 34
209 CGCGCGCTCAGCAAAATCGCGCTATTACCGTGGCGGCAAAAGCGGTA 258
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
34 rpProAspMetSerTyrLeuValAlaGluMetArgSerServalLeu 50
259 CTTCACTCACTGCTGATTGCCGTTGAAGGCAACGACGAAATCGAG..... 303
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

```

51 MetArgGlyMet.....LeuAspGlyHisGluHisValArgCysTh 64
304 .....TTGGAACGCTACGTACCTGAAGCGGTGG 331
64 rCysValProLysCysAlaCysLeuGluProThrLeuProAlaAlaL 81
332 CAAATTTGAGCAGCGAAAGTG..... 354
81 euGluGlnAlaArgSerArgValLeuAspAlaHisAlaGlyArgValPro 97
355 .....CGCGCAACCTGATTCAATC 374
98 GlyLeuArgAlaProSerArgAlaAlaHisArgArgValValLeuThrAl 114
375 AGGTTA...TGGACTCGCTTCGCACCCCTCGCTTCAGCAAAATCCCTG 421
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
114 agLysCysArgPheThrGlnGlyPheProGluProPheGluGlyLeuTrpV 131
422 CGGTAGATCGGAGCGCTTGGCC..... 444
131 alAlaSerAlaGluProGlyAlaHisAspGluPheCysValGlyGlu 147
445 .....ATCTTCGTCAATCGATGACACACAA 470
148 HisCysAspAlaHisLeuSerSerIlePheCysHisAlaAlaSerThrMe 164
471 TCCGCTGGCTGCCGACCT.....ACGGTCATCATCA 502
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
164 tProLeuAlaArgAlaProHisSerProMetValThrPheSerAlaL 181
503 AAGAAGCGCGCGAAGACTTCAAAACGCGGCTGTGTGTTATTGAGCCGCTG 552
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
181 euSerAlaGlyGluAsnTyrLeuArg..... 189
553 ACCGAACGTAAATCATGTGTATAAGCAGCAGCGCA..... 591
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
190 .....LeuValCysArgAlaSerGlyAlaTyrProProVa 201
592 .....GACGTCCCT 601
201 laspThrLeuThrLeuValSerGlnProGlnProGluAspAlaProC 218
602 CTGAAATGCTGCCATATCGAACACATGAATTTGGCGGC..... 642
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
218 ysluThrThrAlaGlyThrAsnAlaAspSerThrGlyHisValGlyMet 234
643 .....CCGCA 647
235 AlaCysValArgSerAspAlaLeuAlaGlyAlaAlaCysAlaValGlnHi 251
648 TCCTGCGCGCTTGAGTGGCAGCAGCATTCATTTCATC..... 684
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
251 saArgGlyValThrThrSerAlaArgIleValLeuValProAlaAsnAspG 268
685 ..GAGCCAGTCGCGCGAATAAACCGTGTGGACCATCAATTATCAAGAC 732
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
268 lYalAlysValGlyAlaTyrAlaAspValAspAlaAspPheTyrAlaAsp 284
733 GTG.....ATTGCTATCGGACGTTT 752
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
285 ValProProLeuProGluProGluSerAspSerLeuAlaValHisAlaLe 301
753 GTCTGTAACAGCGCTCTG.....AATACCGAGCGCGTGG 787
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
301 uPheValAlaGlyAsnThrGluLeuTyrValHisGlyThrAlaAlaGly 317
788 TTGCTTTGGCGGCTGCAAG...TCACAAACCGCGCTCTTCGCTACC 834
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
318 ValProSerAlaSerCysArgCysAspThrArgArgCysThrCysVal.. 333
835 GTTTTGGGTGCGAAGTCTCTCAACTTACCGCGCGCAATGG..... 877
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
334 .....LeuAlaProAlaThrTrpThrAlaG 342

```


541 ProAsnSerAla..... AspSerLysMetSerPheAsp..... 551
 237 AATAGCGGCGATTTCCTGAAGCGCGCAGTAATACTAGCGCGGAT 188
 552 LeuProGluLysGlnAspGlyAlaThrSerProGlyA 564
 187 TCTTTTTCCTCAACAGCACTTGGCTTTTTCGACGGCTTACCTTCC 138
 564 laLeuLeuProAlaSerThrThrSerPhePheThrSerAsnProHis 580
 137 TTGATTTCATCGAGGCGGATGCGGACATATCTTCGCAAGCAAGC 88
 581 AspSerLeuValMetAsnThrLeuThrSerPheSerProLeuAsnG1 597
 87 GACTTCGTAATGGCGCGCTCATAATGACTTGTCTCGGTCGCCC 39
 597 ncluthrAlaValGluAlaProSerArgArgThrAspAspGluillePro 613
 seq_name: pir2:C75580

seq_documentation_block:
 adenine deaminase-related protein - Deinococcus radiodurans (strain RI)
 C:Species: Deinococcus radiodurans
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
 C:Accession: C75580
 R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, R.J.; S.; Smith, H.O.; Venter, J.C.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; M. Science 286, 1571-1577, 1999
 A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans RI.
 A:Reference number: A75250; MUID:20036896
 A:Accession: C75580
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-376 <WHI>
 A:Cross-references: GB:AE001863; GB:AE001825; NID:g6460670; PIDN:AAF12376.1; PID:g6460670
 C:Experimental source: strain RI
 C:Genetics:
 A:Gene: DRA0268
 A:Map position: 2

alignment_scores:
 Quality: 111.50 Length: 290
 Ratio: 0.961 Gaps: 15
 Percent Similarity: 40.000 Percent Identity: 26.897

alignment_block:
 US-09-303-518D-131 x C75580 ..

Align seg 1/1 to: C75580 from: 1 to: 376

639 CGGCGCGCA.....TCTGCGCGCTTGAGTGGCAGCGACA 673
 35 ArgProAlaLeuArgGlyAspSerAlaArgLeuAlaHisAlaHi 51
 674 TTCATTTCAT.....CGACCCAGTCGCGCGAATAAACCGTGTGACC 717
 51 sHisThrHisPheGlyAlaAlaGlyArgThrGly..... 62
 718 ATCAATTATCAAGACGTGATTGTTATCGGAGCTTTGTTCTGTAACAGCGCC 767
 63ThrGlyArgPro 66
 768 TCTGATACGAGCGGTGTTGCCCTTGGCGGCTGCAAGTCAACAAC 817
 67 AlaAlaGlyProAlaGlyGlnSerGlyAlaGlyCluProArgVa 83
 818 CGCG...CCCTTCGCTAC.....CGTT 837
 83 lAlaHisProAlaLeuThrSerGlyGlyArgAlaAspArgAlaGlyArgA 100
 838 TTGGGTGCGAAGTGCTCAACTACCGCGCGGAATTGTTGACGCGGA 887

100 laGlySerAlaAlaGlyAlaArgTyrArgGluilleLeuGly...ProAla 115
 888 CAACCGCGGTGATTTCGCGTTCCGTTATTGAACGGTGGATTGCACAGGCG 937
 116 ArgProArgGlyAspArgProTyrPvalSerArg.....ArgLeuArgAl 130
 938 CGCATGATTATTGGGACGCTA..... 959
 130 aAlaAlaArgLeuAlaThrLeuArgGlyAlaGlyAsnLeuArgGlyArgA 147
 960CCACAATCAGATTTCCGTTAT 980
 147 rgGlySerAlaAlaArgArgArgAspAlaProAla.....ArgTyr 161
 981 CGAAGAGCGCGCAGCAAGAGCTGTTGCGGTGGT..... 1016
 162 ArgArgArgProArg.....AlaArgLeuGlyArgGlyHisLeuAr 175
 1016 1016
 175 gSerAlaArgAlaLeuAlaHisAlaProAspValSerArgProAspArgH 192
 1017TGCGCCGCGAGCGGACAAATACTCATCACCGCACCACTCTCG 1061
 192 isArgAlaCysGlyAlaGlyGlnArgArgProAlaGlyArgArg 208
 1062 CCATTTCTTAAACAACTCTTCAAGTTCACGACGCGCTCAACGGCG 1111
 209 ProLeuArgAlaArgValValGluLeuLeuAspValGlyGlnArgPr 225
 1112 GCGACCGCGCATGTACCGATCGGCATCTATCAGCGCGTAATGCGGTG 1161
 225 oAlaArgArgHisProGlyHisGlnHisSer...AlaArgArgAlaSerG 241
 1162 GACATCTCGCTACCTT.....GCTTTT 1184
 241 lyGlyProAlaArgArgGlnArgGlyProAlaArgGlyGlySerGly 257
 1185 GCGCGATTAACTCGTCGCGATACCGACAGCGCGAGCTTTGGTGTGCT 1234
 258 AlaArgThrAlaArgArgHisArgPro..... 267
 1235 TGGATTGGACGAAGACCTCGCTTGTGCAGCTTCGTCGCCGCGGC 1284
 268 .GlyArgGlyArgGlyProArgAlaValAlaProAlaLeuArgGlys 284
 1285 AATACGAATACGCGCGCT 1304
 284 er...AspAspArgProAla 289

seq_name: pir2:JC4176

seq_documentation_block:

pyruvate,water dikinase (EC 2.7.9.2) - Pyrococcus furiosus

N:Alternate names: phosphoenolpyruvate synthetase

C:Species: Pyrococcus furiosus

C:Date: 10-Sep-1995 #sequence_revision 27-Oct-1995 #text_change 21-Jul-2000

C:Accession: JC4176

R:Jones, C.E.; Fleming, T.M.; Piper, P.W.; Littlechild, J.A.; Cowan, D.A.

Gene 160, 101-103, 1995

A:Title: Cloning and sequencing of a gene from the archaean Pyrococcus furiosus with

A:Reference number: JC4176; MUID:95354939

A:Molecule type: DNA

A:Residues: 1-817 <JON>

A:Cross-references: EMBL:X80819; NID:g967059; PIDN:CAA56785.1; PID:g967060

A:Experimental source: Vcl DSM 3638

C:Genetics:

A:Gene: ppsA

C:Function:

A:Description: catalyzes the ATP-dependent conversion of pyruvate into phosphoenolpyr

A:Pathway: gluconeogenesis

C;Superfamily: Escherichia coli pyruvate,water dikinase; phosphotransferase system enzym
C;Keywords: phosphoprotein; transferase
F;66-816/Domain: phosphotransferase system enzyme I homology <PT1>

alignment_scores:
Quality: 111.00 Length: 489
Ratio: 0.459 Gaps: 24
Percent Similarity: 49.489 Percent Identity: 21.268

alignment_block:
US-09-303-518d-131 x JC4176 ..

Align seg 1/1 to: JC4176 from: 1 to: 817

```

31  CCATCGCGGAGAGCGGAGCAAGTCAATTTATGAC..... 66
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
235  ProValThrAsnAsnArgAsnGluMetIleAsnAlaSerTirpGlyLe 251
67  .GCGCGCGCATACCGAAGTCGCGTGTGGCGAAGATATGCGGCA 115
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
251  uGlyGluAlaValValSerGlyAlaValThrProAspGluTyrIleValG 268
116  TGGCGCCCTCGATGAAATCAAGAGGTGAAGCCGCTCAAAAGGCCAA 165
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
268  LuLysGlyThrTrpLysIleLysGluLysValIleAlaLysGluVal 284
166  GTGCTGTTTGAACAAAAAAGATCGG.....GCGCTAGTATT 203
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
285  MetValIle.....ArgAsnProGluThrGlyArgGlyThrValMe 298
204  TACTGCGCGGCTTCAGGCAAAATCCCGCTATTACCGCTGC..... 246
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
298  tVal.....LysValAlaGlu...TyrLeuGlyProGluTr 309
247  ....GAAAGCGCGTACTT.....CAGTCAGTCGTGATTCGCGTGAA 285
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
309  rpValGluLysGlnValLeuThrAspGluGlnIleIleGluValAlaLys 325
286  GGAACAGCAAAATGAGTTCGACGCTACCTACCTGAAGCGCTGCA... 333
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
326  MetGlyGlnLysIleGluAspHisTyrGlyTrpProGlnAspIleGluTr 342
334  AAATTGAGCAGCAAAAGTGGCGCGCAACCTGATTCATCAATCAGCTTAT 382
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
342  pAlaTyrAspLysAspGlyLysLeuTyrIleValGlnSerArgProI 359
383  GGACTGCGCTTCGACCCGTCGCTTCAGCAAAATCCCTGCGGTAGATGCC 432
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
359  leThrThrLeuLysGluGluAlaThrAlaGlu...GluAlaGluGluVal 374
433  GAGCGGTGCGCATCTTCGTCAATGCGATGGACACCAATCCG...CTGGC 479
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
375  GluGluAlaGluValIleLeuLysGlyLeuGlyAlaSerProGlyIleGI 391
480  TGCGGACCCCTACGCTCATCATCAAGAACGCCGCGAA.....GACTTCA 523
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
391  yAlaGlyArgValValValIlePheAspAlaSerGluIleAspLysValL 408
524  AACCGCGC...CTGTGCTATTGAGCGCGCTGACCGAACGTAATCCAT 570
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
408  ysGluGlyAspIleLeuValThrMetThrAsnProAspMetValPro 424
571  GTGTGTAAAGCAGCGGCGGAGCTGCGCTCTGAAATGCT..... 612
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
425  AlaMetLysArgAlaAlaIleValThrAspGluGlyGlyArgThrSe 441
613  ....GCCAATATCAACACATGAATTTGGCGGCCCGCATCTGCGCGCT 658
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
441  rHisAlaAlaIleValSerArgGluLeuGlyIlePro..... 453
659  TGAGTGGCACCCATTCATTCATCGAGCCAGCTGCGCGCGCAATAAACCC 708

```

```

453  ..... 453
709  GTGTGGACCAATCAATATCAAGACGTAATGCTATCGACGTTGTCGT 758
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
454  .....CysValValGlyThrLysGluAl 461
759  AACAGCGCGTCTGAATACCGAGCGGTGCTTGGCGCGCTGCA... 807
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
461  aThrLysLysLeuLysThrGlyMetTyrValThrValAspGlyThrArg 478
808  ....GTCACAAACCGCGCTCTGCGTACCGTGTGGTGGTGGCAAGGTG 852
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
478  lyLeuValTyrLysGlyIleValLysSerLeuValLysLysLysGlu 494
853  TCTCAACTTACCGCGCGGATGTTGACGCGACAAACCGCTGATTTC 902
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
495  AlalysAlaGluGlyGlnValValAlaGlyAlaProLeuValTh 511
903  CGGTTCG.....GTATTGAACGGTG 922
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
511  rGlyThrMetValLysValAsnValSerMetProGluValAlaGluArg 528
923  CGATTGCAACAGCGCG.....CATGAT 945
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
528  laAlaAlaThrGlyAlaAspGlyValGlyLeuLeuArgAlaGluHisMet 544
946  TATTGGGACGCTACCAATACAGATTTCGTTATCGAAGAGCGCGCAG 995
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
545  lleLeuSerIleGlyGlnHisProIleLysPheIleLysGluGlyLysGI 561
996  CAAAGAGCTGTTGCGGTGGTGGCGCGACCGCAAAATATCTCCATCA 1045
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
561  uGluLeuValGluLysLeuAlaGluGlyIleGluLysValAlaAla... 577
1046  CGCGCACCATCTCGCGCATTT.....CCTAAAA 1074
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
578  .AlaPheTyrProArgProValTrpTyrArgThrLeuAspAlaProThr 594
1075  AACAACTCTTCAAGTT.....CACGACACCGCTCAACGGCGCGCA 1115
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
594  snGluPheArgGluMetProGlyGlyGluAspGluProGluGluArgAsn 610
1116  CCGCGCATGTACCGCTCGCACCTATGACGCGTAAATCGCTTGACA 1165
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
611  ProMetLeuGlyTrpArgGly.....IleArgArgGlyLeuAspGI 624
1166  TCCTCGCTACCTTCTTTGGCGGATTTAATCGTCGCGATACCCACAGC 1215
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
624  nProGluLeuLeuArg...AlaGluPheLysAlaIleLysLysValValG 640
1216  GCGCGAGCTTTGGTGTGATTTGGCAAGACCTCGCTTTGTG 1265
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
640  LuLysGlyTyrAsnAsnIleGly.....Val 648
1266  CAGCTTCGCTCG.....CCGCGCAAAATACGAATACGCGCGCTGT 1306
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
649  MetLeuProLeuValSerHisProGluGlnIleArgGluAlaLysArgII 665
1307  TGCAGCAAGTGTGGA 1322
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
665  eAlaArgGluValGly 670
seq_name: pir2:T45134

```

seq_documentation_block:
Hypothetical protein [imported] - Microbacterium ammoniaphilum (fragment)
C;Species: Microbacterium ammoniaphilum
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C;Accession: T45134
R;Striebel, H.M.; Seeber, S.; Jarsch, M.; Kessler, C.
Gene 172, 41-46, 1996
A;Title: Cloning and characterization of the MamI restriction-modification system fro
A;Reference number: 222923; MUID:96257250

A:Accession: T45134
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-529 <STR>
 A:Cross-references: EMBL:X79027; NID:g984667; PIDN:CAA55649.1; PID:g1679831
 A:Experimental source: ATCC 15354

alignment_scores:

Quality: 110.50 Length: 490
 Ratio: 0.531 Gaps: 25
 Percent Similarity: 42.449 Percent Identity: 24.694

alignment_block:

US-09-303-518d-131 x T45134 ..

Align seg 1/1 to: T45134 from: 1 to: 529

```

18 AGGTCTAAATCTGCCATCGCGG..... 41
   |||:|||||
99 ArgAspArgAlaGluArgGlyArgGluProAlaArgProAlaAl 115:
42 .CAGACCGGAGCAATGTCGCGCATCGCCCTCGATGAAATCAAGGA 140
   |||:|||||
115 aArgAspAlaArgArgHisLeuArgLeuGlyAlaHisProAlaArgG 132
   |||:|||||
91 TTGCTGGCGAAGATATGTCGCGCATCGCCCTCGATGAAATCAAGGA 140
   |||:|||||
132 lyAlaLeuArgArgHisArgHisArgAlaArgGlyGlu...ArgGly 147
141 AGGTGAAGCGCTCAAAAGCCCAAGTCTGTTGAAGACAAAAGAATC 190
   |||:|||||
148 ArgGlyProArgGlnGlnValPro.....ArgGlnHisProAr 160
191 CGGCGCTAGTATTACTCGCGCGCTTCAGGCAAAATCGCGCTATTTCAC 240
   |||:|||||
160 gGly.....ArgArgAspArgAlaGlyArgP 169
241 CGTGGCGAAAGCGCTACTTCAGTCTGCTGATTCGCGTGAAGCCAA 290
   |||:|||||
169 roGlyLeuHisArgArg.....ArgArgAlaArgArgGln 182
291 CGACGAATCGAGTTCGACGCTACGTACCTGAAGCGCTGGCAAAATTCGA 340
   |||:|||||
183 GlyArgProGlnValArg.....HisGlyAspAspG 193
341 GCACGAAAGTGGCGCGCAACTGATTCAATCAGGCTTATGACCTGCG 390
   |||:|||||
193 nHisArgAlaAspProArgArgProArgAspProArg.....AlaG 207
391 CTTGCGCACCGCTCGTTCAGCAAAATCCCTGCGTAGATCGCGACCGTT 440
   |||:|||||
207 lyHisHisProLeuArgGluGlyGlnGlyGlnGluAlaArgProLeu 223
441 CGCATCTTCGTCATTCGATGACACCAATCCGCTGGCTGCGACCCCTA 490
   |||:|||||
224 Arg.....AlaArgGlyGlnGlyGlyProGlyGlyProAl 236
491 CGGTCTCATCT.....CAAAGACCGCCCGAAGATTC 522
   |||:|||||
236 aGlyArgHisProAlaLeuGlyAspGluAspArgProArgAlaArgGlnG 253
523 AAACGGGCGCTGTTGTTATGAGCGCGCTACCGCAACGTAAATCCATGT 572
   |||:|||||
253 lyAlaArg..... 255
573 GTCTAAGCAGCGCGCAGAGCTGCGCTCTGAAATCTGCCAATATCG 622
   |||:|||||
256 .....AlaAlaHisArgGluAlaLeuHisLeuArgLeuGlnArgAr 269
623 AAACACATGAATTTGGCGCGCGCATCTCGCGCTTGAAGTGGACCGAC 672
   |||:|||||
269 gArg.....GlyArgAlaAspGlyCysArgAlaGlnGlyArgAla 283

```

```

673 ATTCAATTCATCGACCCAGT.....CGCGCGCAATAAAC 707
   |||:|||||
283 rgArgAlaArgAlaGlyGlyGlyGlyLeuProArgArgGlu..... 297
708 CGTGTGGACCATCAATTCAAGACGCTGATTGCTATCGGACGTTTGTTCG 757
   |||:|||||
298 .....AspArgValArgAlaHisArgProArgProAr 308
758 TAACAGGCGCTCTGAATAC.....CGAGCGCTGTTGCTTGGCGG 798
   |||:|||||
308 gGlyArgGlyGlyAlaAlaArgLeuAspArgAlaGlyGlyValGlyAla 325
799 GGCCTGCA..... 806
   |||:|||||
325 rgProAlaArgProArgArgLeuArgHisProArgProAlaAspLeuPro 341
807 .....AGTCAACAAACGCGCCTCTTCGCTACCGTTTGGGTGCGGA 847
   |||:|||||
342 HisGlyGlyProGlnGlyValAlaArgLeuAspHisProGlnGlyLeuG 358
848 AGGTGCTCAACTTACCGCGCGCAATTGCTGACGCGGACACCGCGTG 897
   |||:|||||
358 uGlyAlaAlaGlySerArgArg.....HisProHisA 369
898 ATTTCGCGTTCGCTATTGAA.....CGGTGCGATTGCACAGCGCGCA 941
   |||:|||||
369 rgLeuArgAlaArgLeuHisGlnGlyArgGlyAspLeuLeuArgArg... 384
942 TGATTATTGGGACGCTACCAACATCAGATTTCGTTATCGAAGAAGCC 991
   |||:|||||
385 .....ProArgArgAspArgLeuGlyArgArgGlyPr 395
992 GCACAAAGACGCTTCGCGCTGCTGCGCGCGCGCGGACGCGCAACTCC 1041
   |||:|||||
395 oArgGlnGlyGln.....GlyAlaHisGlyGlyGlnGlyLeuA 408
1042 ATCAGCGCGCACCACTCTCGCGCCTTCCTTAAATAACAACTCTCAAGTT 1091
   |||:|||||
408 rgHisAlaGlyArgArgArgArgGlyValProSerGlnLeuAlaValL 424
1092 CACGACAG.....CGTCAACGCGCGCGCGCGCGCATGGTACCGA 1132
   |||:|||||
424 euArgGlnGlyValCysGlnAlaSerAlaThrLeuAlaLeuTrpMetThr 440
1133 TCGSCATTTATGAGCGCGTAATCCGCTTGGACATCTCTGCTTACCTGTT 1182
   |||:|||||
441 SerGlyArgLeu.....LeuArgGluLeuAlaGlyLeuLeuCysAr 454
1183 TTGCGCGATTAAATCTCGCGCGCATACCGACA..... 1213
   |||:|||||
454 gAspSerValGluSerHisLeuAspValThrThrTrpArgAlaAspAla 471
1214 .....CGCGCGAGCTTTGGTGTGTTGGAATTTGGAAGAAAGACCTCG 1258
   |||:|||||
471 erileSerArgArgThrThrTrpValMetAsn...ThrArgLeuArgGln 486
1259 CTTTGTGCGAGCTCTGCTCGCGCGCAATACGAATACGGCCGCTGTTG 1308
   |||:|||||
487 AlaGlnLysSerLeuAlaAlaArgArgProThrThrSerSerIleSe 503
1309 CGCAAAAGTCTGGAAA 1324
   |||:|||||
503 rArgSerCysTrpGln 508

```

seq_name: pir2:T34434

seq_documentation_block:

hypochemical protein K06A9.1a - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
 C:Accession: T34434
 R:Geisel, C.; Gattung, S.

A:Title: DNA sequence analysis of the photosynthesis region of Rhodobacter sphaeroides 2
 A:Reference number: Z25222; MUID:20115911
 A:Accession: T50711
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-569 <CHO>
 A:Cross-references: EMBL:AF195122; PIDN:AAF24255.1
 A:Experimental source: strain 2.4.1
 C:Genetics:
 A:Gene: ureC
 C:Superfamily: urease 62K chain; urease 62K chain homology
 C:Keywords: hydrolase

alignment_scores:
 Quality: 108.00 Length: 373
 Ratio: 0.643 Gaps: 20
 Percent Similarity: 45.040 Percent Identity: 22.788

alignment_block:

US-09-303-518d-131 x T50711 ..

Align seg 1/1 to: T50711 from: 1 to: 569.

```

121 CCCTCGATGAATCAAGGAGGTGAAGCGCTCAAAAAGGCCAAGTGT 170
|||||
116 ProGlyThrGluIleAlaGlyGlu.....GlyArgIleLe 128
171 G.....T 172
128 uThAlaGlyGlyMetAspAlaHisIleHisPheIleCysProGlnGlnI 145
173 TTGAAGACAAAGAATCCGGGGTAGTATTT.....ACT 207
|||||
145 LeGluAspSerLeuHisSerGlyIleThrMetLeuGlyGlyGlyThr 161
208 CGCGCGGCTTCAGGCAAAATCGCGCT.....AT 236
|||||
162 GlyProAlaHisGlyThrLeuAlaThrThrLeuHisAlaGlyAlaLeuAl 178
237 TCACCGT.....GGCGAAAGCGCGTACTTCAGTCAGTCGTGA 274
|||||
178 aHisArgAlaAspAlaAlaGlyGlyArgArgLeuProAspGlnProArgL 195
275 TTGCGGTTGAAGCAGCAGCAATCGAGTTCGAA..... 309
|||||
195 euArgGlyGlnGlyGlnCysGlnProAlaArgAlaArgGlyAlaGly 211
310 .....CGCTACGTACTGAAGCGCTGGCAAAATTCAGCAGCGGAAA 350
|||||
212 ProArgGlyArgLeuValProGluAlaAlaArgGlyLeuGlyHisAl 228
351 AGTGGCGCGCACTGATTCATTCAGGCTT.....ATGGACTCGGCTT 393
|||||
228 aCysGlyHisArgLeuProValArgArgProHisGlyCysAlaG 245
394 CG.....CACCGTCGCTCAGCAAAATCCCTGCGTAGATGCCGA 434
|||||
245 lYAspAspProHisArgHisAlaGluArgGluArgArgGluHis 261
435 GCGGTCGCGCATCTCGT.....CAATGCGATGGACCAATCCGC 475
|||||
262 AlaGlyArgHisArgAlaHisAspProCysValProHisArgGlyCy 278
476 TGGCTCGCGCCTACGCTCATCAAGAAGCGCGCAAGACTTCAA 525
|||||
278 sGlyArgProCysAlaArgHisGlnGlyGlyArgArgGlu 295
526 CCGCGCGCTGT.....GGTATTGAGCGCGCTGACCGAAGCTA 562
|||||
295 rHisProValLeuAspGlnSerHisHisAlaLeuHisArgGlnTyrGly 311
537 .....
```

```

312 GlyGlyAlaSerArgHisAlaAspGlyLeu...ProProSerArgProLe 327
563 AAATCCATGTGTGTAAAGCAGCGCAGACGTGCCGTCTGAAATGCT 612
|||||
327 uAspPro.....ArgGlyC 332
613 GCCAATATCGAAACACATGAATTTGGCGGCCGCATCTGCCGGCTTGAG 662
|||||
332 ysGlyLeuArgArg.Glu.....ProHisProAlaArgLysTh 344
663 TGGCAGCGCACATTTCATTCAGCAGCCAGTCGGCGCGAATAAAACCGTGT 712
|||||
344 rIleAlaAlaGluAspIleLeuHisAspMetGlyAla.....P 357
713 GGACCATCAATTACAGACGTGATTGCTATCGGACGCTTTGTCGTAACA 762
|||||
357 heSerValIleSerSerAspSerGlnAlaMetGlyArgVal..... 370
763 GGCCGCTCTGAATACCGAGCGCGTGTTCCTTGGCGGCCCTGCAAGTCAA 812
|||||
371 GlyGluValIleThrArgThrTrpGlnThrAlaHisLysMetLysValGI 387
813 CAAACCGCGCCTCTTGGTACCGCTTTTGGTGGGAAGGTCTCTCAACTTA 862
|||||
387 nArgGlyArgLeuAlaGluGluThr...GlyAlaAsnAspAsnGlnArgV 403
863 CCGCGCGCGAATTGTTGACGCGGACAAACCGCGTGTTCGCGTTCGGTA 912
|||||
403 al.....ArgArgTyrIleAlaLysTyrThr 411
913 TTGAACCGTGGGATTGCACAAAGCGCGCATGATTATTGGGACGCTPACA 962
|||||
412 lIleAsnProAlaIleAlaHisGlyLeuSerArgHisIleGlySer..... 426
963 CAATCAGATTTCGTTATCGAAGAAGCGCGCAGCAAAAGAGCTGTCGGCT 1012
|||||
427 .....ValGluGluGlyLysArgAlaAspLeuValLeuT 438
1013 GGGTTCGCGCGCAGCG 1029
438 rp.....GlnPro 440
```

